МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
"ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"

О. Б. Білоцерківський, О. О. Замула, Н. В. Ширяєва

СТАТИСТИКА

Текст лекцій

для студентів спеціальностей
7.050206 "Менеджмент зовнішньоекономічної діяльності"
і 6.030508 "Фінанси"

Затверджено
редакційно-видавничою
радою університету,
протокол № 2 від 18.09.2009

Харків НТУ „ХПІ” 2009

ББК 22.213
Б 78

УДК 534.1

Рецензенти: *В.А. Смоляк*, канд. екон. наук, доц., ХНЕУ;
Т.В. Данько, канд. екон. наук, доц., НТУ "ХПІ"

Білоцерківський О. Б. Статистика: текст лекцій
Б 78 О. Б. Білоцерківський, О. О. Замула, Н. В. Ширяєва. – Харків : НТУ "ХПІ", 2009. – 96 с.

Текст лекцій містить основи загальної теорії статистики, включаючи групування статистичних даних, абсолютні, відносні і середні величини, статистичні розподіли, вибіркове спостереження, ряди динаміки, індекси та їх використання в економіко-статистичних дослідженнях. Представлено типові приклади з розв’язаннями за матеріалом, що вивчається.

Призначено для студентів спеціальностей 7.050206 "Менеджмент зовнішньоекономічної діяльності" та 6.030508 "Фінанси".

Іл. 4. Табл. 15. Бібліогр. 5 назв.

ББК 22.213

ISBN

ЗМІСТ

ВСТУП

Статистика здійснює збирання, обробку та аналіз даних про масові соціально-економічні явища, які характеризують всі сторони життя та діяльності населення, виявляє взаємозв'язки різних сторін в економіці, вивчає динаміку її розвитку та прийняття ефективних управлінських рішень на всіх рівнях, що складає *предмет* дисципліни.

Її метою є оволодіння основами статистичного вимірювання, методами узагальнення та аналізу інформації про соціально-економічні явища та процеси, про закономірності суспільного життя.

Після вивчення статистики студенти повинні *знати* можливості статистичних методів спостереження, зведення та групування статистичних даних, економічну суть статистичних показників, методи аналізу конкретних явищ і процесів суспільного життя. Студенти повинні *вміти* проводити статистичну обробку даних з побудовою статистичних таблиць і графіків, рядів розподілу, аналізувати результати і робити науково обґрунтовані висновки.

У даному тексті лекцій викладаються теоретичні основи курсу «Статистика», включаючи групування статистичних даних, абсолютні, відносні і середні величини, статистичні розподіли, вибіркове спостереження, ряди динаміки, індекси та їх використання в економіко-статистичних дослідженнях. Текст лекцій містить достатню кількість розв’язаних задач.

Даний текст лекцій розрахований на студентів спеціальностей 7.080206 "Менеджмент зовнішньоекономічної діяльності" та 6.030508 "Фінанси", що вивчають курс "Статистика". Він також буде корисним для студентів інших економічних спеціальностей.

1. ПРЕДМЕТ І МЕТОД СТАТИСТИКИ

1.1. Статистика як наука

Термін „статистика” визначається сукупністю латинських та італійських слів: *"status"* (становище, стан справ); *"stato"* (керована область, держава); *"statista"* (державний чоловік, політик, знавець держави).

Статистика має багатовікову *історію.*

Найбільш ранні відомості були про облік чисельності населення у Стародавньому Китаї (ХХІІ ст. до н.е.), потім Стародавньому Єгипті (вимір та оцінка земель), Стародавній Греції (чисельність і майнове положення класів населення). Достатньо досконалі форми організація статистики набула у Стародавньому Римі (VI ст. до н.е.) у вигляді цензу (даних за кожним римським громадянином про його ім'я, стать, вік, майнове положення тощо), цензи повторялись через кожні 5 років.

У наукове користування термін „статистика” було введено німецьким вченим, професором Геттингенського університету Г. Ахенвалем у 1743 році для визначення сукупності знань, які характеризують державний устрій, добробут країни. Однак таке визначення далеке від сучасного тлумачення поняття „статистика”.

Статистика – це наука, яка вивчає розміри і кількісні співвідношення масових суспільно-економічних явищ і процесів у нерозривному зв'язку з їх якісним змістом.

Основою для вивчення масових явищ є *закон великих чисел.* Сутність його полягає в тому, що кожне одиничне явище випадкове (воно може бути або не бути), але у з'єднанні великої кількості таких явищ випадковість зникає.

Об'єкти статистичного аналізу – найрізноманітніші явища і процеси суспільного життя.

Основні *відмінності* від інших суспільних наук такі. По-перше, статистика вивчає не поодинокі, а масові соціально-економічні явища і процеси суспільного життя. По-друге, предметом статистики є кількісна сторона явищ і процесів суспільного життя.

Кількісну сторону масових суспільних явищ і процесів статистика виражає у вигляді статистичних показників (чисел).

Статистичним показником називають узагальнену числову характеристику будь-якого масового явища (процесу) з його якісною визначеністю в конкретних умовах місця та часу. (Прикладами статистичних показників є кількість працюючих на підприємстві на початок року, обсяги виробленої та реалізованої продукції, собівартість, рентабельність виробництва тощо).

Статистичні показники можуть бути виражені у вигляді *абсолютних* і *відносних* величин.

Якщо статистичний показник стосується окремого явища (наприклад, конкретного підприємства), то його називають *індивідуальним,* якщо ж сукупності явищ (наприклад, однотипних підприємств регіону), то *узагальненим,* або *зведеним.*

Статистика оперує з відповідними *категоріями,* тобто поняттями, які виражають суттєві, всебічні властивості явищ дійсності. До основних категорій статистики можуть бути віднесені:

а) статистична закономірність;

б) статистична сукупність;

в) одиниця сукупності;

г) ознака сукупності;

д) варіація ознаки;

е) показник.

Статистична закономірність – це повторюваність, послідовність і порядок у масових соціально-економічних явищах (процесах).

Статистична сукупність – це множина одиниць (об'єктів, явищ), що об'єднуються однією якісною основою, але відрізняються між собою за рядом ознак, яка піддається статистичному вивченню.

Окремі елементи статистичної сукупності називають *одиницями сукупності.*

Одиниці сукупності характеризуються однією або кількома ознаками. *Ознака* – це статистичний еквівалент властивостей одиниць сукупності. (Так, для одиниці статистичної сукупності „підприємство" ознаками можуть бути: обсяги виробленої продукції, співвідношення власних та запозичених коштів, чисельність робітників тощо).

1.2. Класифікація ознак

Однією з особливостей статистичної сукупності є наявність *варіацій* ознак, тобто відмінностей, коливань у числових значеннях окремих одиниць сукупності. Ознаки, які набувають різних значень, називають *варійованими*. (Прикладами *варійованих* ознак людини є вік, стать, сімейний стан, рівень освіти, а підприємства – спеціалізація, форма власності, рентабельність виробництва тощо; неварійованими – дата народження).

Варійовані ознаки поділяють на кількісні та атрибутивні (якісні). *Кількісні* ознаки виражаються числами (урожайність, заробітна плата, продуктивність праці та ін.). *Атрибутивними* називають ознаки, які не підлягають числовому вираженню і характеризують словами описові риси (професія, галузь та ін.).

За характером варіювання кількісні ознаки поділяють на дискретні та неперервні. *Дискретними* називають такі кількісні ознаки, які можуть набувати тільки цілочислових значень (кількість автомобілів, кількість членів сім'ї та ін.). *Неперервними* кількісними ознаками є такі, які можуть в окремих межах набувати будь-яких значень (вік людини, стаж роботи, собівартість продукції тощо).

За важливістю ознаки поділяються також на істотні (основні) та неістотні (другорядні). *Істотними* називають такі ознаки, які є головними для даного явища. (Наприклад, для підприємства ними є обсяг виробленої та реалізованої продукції, кількість працівників, продуктивність праці та ін.) *Неістотними* є такі ознаки, які не пов'язані безпосередньо з суттю досліджуваного явища, наприклад: підпорядкування підприємства, територіальна належність тощо.

Ознаки, що характеризують статистичну сукупність, взаємопов'язані між собою, тому розрізняються факторні та результативні ознаки. *Факторні* ознаки – це незалежні ознаки, які впливають на інші ознаки і є причиною їх зміни. *Результативними* ознаками називають залежні ознаки, які змінюються під впливом факторних ознак (табл.1.1). (Так, кваліфікація, стаж роботи – факторні ознаки; продуктивність праці – результативна ознака).

Таблиця 1.1 – Класифікація ознак

Сторони ознаки	Класифікація
наявність варіацій	варійовані, неварійовані
спосіб представлення	кількісні, атрибутивні (для варійованих)
характер варіювання	дискретні, неперервні (для кількісних)
важливість	істотні, неістотні
залежність	факторні, результативні

Ознаки мають різний рівень вимірювання, що відображається у різних видах *шкал.* Існує така *класифікація* шкал ознак: *номінальна,* яка встановлює шкалу найменувань; *порядкова,* яка встановлює відношення подібності і послідовності; *матрична,* де за допомогою звичайних чисел вимірюються явища, ресурси, результати господарсько-фінансової діяльності.

1.3. Методологія статистики

Для вивчення свого предмету статистика розробляє і використовує різні методи, сукупність яких утворює *статистичну методологію.*

Теоретичною основою статистики як суспільної науки є філософія, математика та економічна теорія (політична економія, макро- і мікроекономіка).

До методів належать: статистичне спостереження; зведення і групування даних; визначення абсолютних, відносних та середніх величин, показників варіації, динаміки; використання вибіркового методу, кореляційно-регресивного аналізу, табличного і графічного методів тощо.

Будь-яке статистичне дослідження складається з трьох послідовних етапів:

1) статистичне спостереження;
2) зведення, класифікація та групування статистичних даних;
3) аналіз статистичної інформації.

Розглянемо кожен етап статистичного дослідження більш докладно.

На *першому етапі* використовується *метод масового статистичного спостереження,* який забезпечує всебічність, повноту та репрезентативність початкової інформації.

На *другому етапі* зібрана в ході масового спостереження інформація підлягає обробці *методом зведення, класифікацій та статистичного групування.* При обробці статистичної інформації обчислюються абсолютні, відносні, середні величини, статистичні коефіцієнти тощо.

На *третьому етапі* проводиться аналіз статистичної інформації з використанням *узагальнених статистичних показників:* абсолютних, відносних та середніх величин; варіацій; параметрів тісноти зв'язку та швидкості зміни соціально-економічних явищ за часом, індексів тощо.

Проведення аналізу дозволяє перевірити причинно-наслідкові зв'язки явищ і процесів, визначити вплив та взаємодію різних факторів, оцінити ефективність прийнятих управлінських рішень, можливі економічні та соціальні наслідки створюваних ситуацій.

Для аналізу статистичної інформації найчастіше використовують *табличні та графічні методи.*

Контрольні запитання

1. Статистика як наука, статистичні показники та їх види, категорії статистики.
2. Класифікація ознак.
3. Поняття статистичної методології.
4. Етапи статистичного дослідження.

2. СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ

2.1. Сутність статистичного спостереження та вимоги до нього

Статистичне спостереження є першим етапом статистичного дослідження суспільних явищ і процесів.

Джерелами статистичного спостереження є соціально-економічні явища, які досліджуються для подальшого аналізу.

Статистичне спостереження здійснюється шляхом *реєстрації* (запису) відповідних ознак явищ і процесів в облікових документах для подальшого узагальнення.

Статистичне спостереження повинно задовольняти таким *вимогам:*

- мати конкретне значення, наукову та практичну цінність;
- забезпечувати збирання масових даних;
- бути орієнтованим на збирання даних, які безпосередньо характеризують об'єкт вивчення;
- ураховувати факти і події, під впливом яких здійснюється зміна стану об'єкта;
- забезпечувати достовірність інформації, яку збирають, для чого здійснюється ретельна перевірка якості зібраних даних;
- проводитися на науковій основі за заздалегідь розробленим планом.

2.2. Форми, види та способи спостереження

Форми спостереження

Статистичні дані можна отримати різними шляхами. З організаційної точки зору розрізняють *три форми* статистичного спостереження: звітність; спеціально організоване статистичне спостереження; реєстри (рис.2.1).

Статистична звітність – це така форма спостереження, коли кожний суб'єкт діяльності (підприємство, організація, установа) подає свої дані у статистичні органи (характерна для Україні). Дані подаються у вигляді звітів, які підводять результати роботи суб'єкта діяльності за звітний період.

Зміст звітності визначається її *програмою.* Звітність здійснюється за встановленою формою і називається *табелем звітності.* Тут наводяться дані про узагальнені статистичні показники, (наприклад: фонд місячної заробітної плати тощо)

Розрізняють типову та спеціальну форму звітності. *Типова звітність* має однакову форму і зміст для всіх суб'єктів діяльності. *Спеціальна звітність* виражає специфічні моменти для окремих підприємств.

За принципом періодичності звітність поділяють на *річну* та *поточну.* Остання включає такі види звітності як квартальну, місячну, двотижневу, тижневу.

Залежно від способу передачі інформації розрізняють *поштову, телеграфну, факс-модемну* звітність.

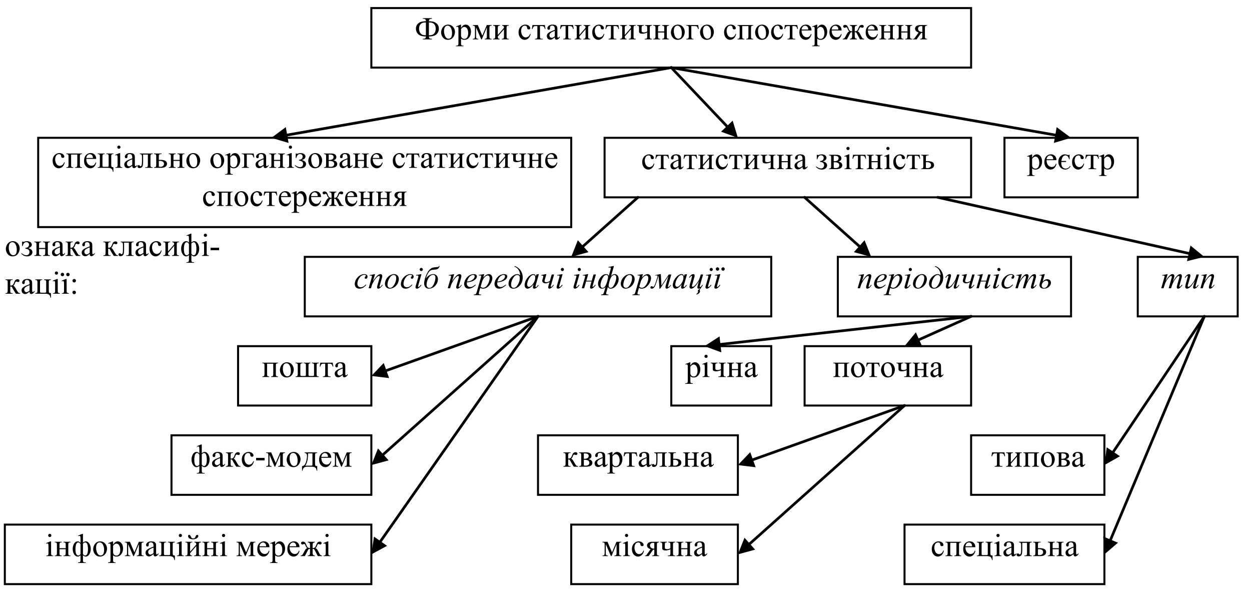

Рисунок 2.1

Спеціально організоване статистичне спостереження є другою формою звітності і має на меті отримати відомості, які не охоплені звітністю (переписи, обліки, спеціальні обстеження, опитування).

Ще одною формою обстеження є *реєстр* – перелік одиниць об'єкта спостереження із зазначенням ознак, який складається та оновлюється під час постійного обстеження. (Наприклад, реєстр населення – поіменний перелік мешканців регіону, який регулярно переглядається і містить паспортні та податкові відомості про кожного мешканця).

Види спостереження

Статистичне спостереження розрізняється залежно від часу реєстрації даних та ступеня охоплення одиниць спостереження (рис.2.2).

Спостереження *за часом реєстрації даних* поділяються на поточне, періодичне та одноразове. При *поточному спостереженні* звітність постійно реєструється по мірі виникнення даних. *Періодичне спостереження* проводиться через певні проміжки часу, наприклад: перепис населення, виробничих площ, технологій, а також обстеження суб'єктів бізнесу щодо можливості інвестування.

За ступенем охоплення одиниць спостереження буває суцільним та несуцільним. *Суцільним* називають таке спостереження, при якому обстежуються всі без винятку одиниці сукупності, наприклад: перепис населення; облік випуску продукції та ін. При *несуцільному* спостереженні обстежується тільки частина одиниць сукупності.

Несуцільне спостереження у свою чергу поділяють на вибіркове, монографічне, основного масиву, анкетне, моніторинг.

Вибірковим називають таке спостереження, при якому обстеженню підлягає певна частина одиниць сукупності, яку отримали на основі випадкового відбору; цей вид статистичного спостереження отримав велике визначення в статистичній практиці.

Монографічне спостереження характеризується тим, що здійснюється докладне і всебічне обстеження окремих одиниць досліджуваної сукупності (опис нових технологій, виробництва окремих видів продукції, передового досвіду тощо).

Спостереження основного масиву – це спостереження за частиною найбільших одиниць, питома вага яких переважає в загальному обсязі сукупності (дослідження найбільш крупних транспортних вузлів у загальній структурі вантажного потоку; спостереження за торгівлею на ринках у місцях, де мешкає більшість міського населення та ін.).

Анкетне спостереження ґрунтується на розсиланні анкет певному колу осіб або установ.

Моніторинг є різновидом несуцільного *спостереження* за станом певного середовища (наприклад, моніторинг стану здоров'я мешканців зони посиленого екологічного контролю, моніторинг валютних торгів та аукціонів тощо).

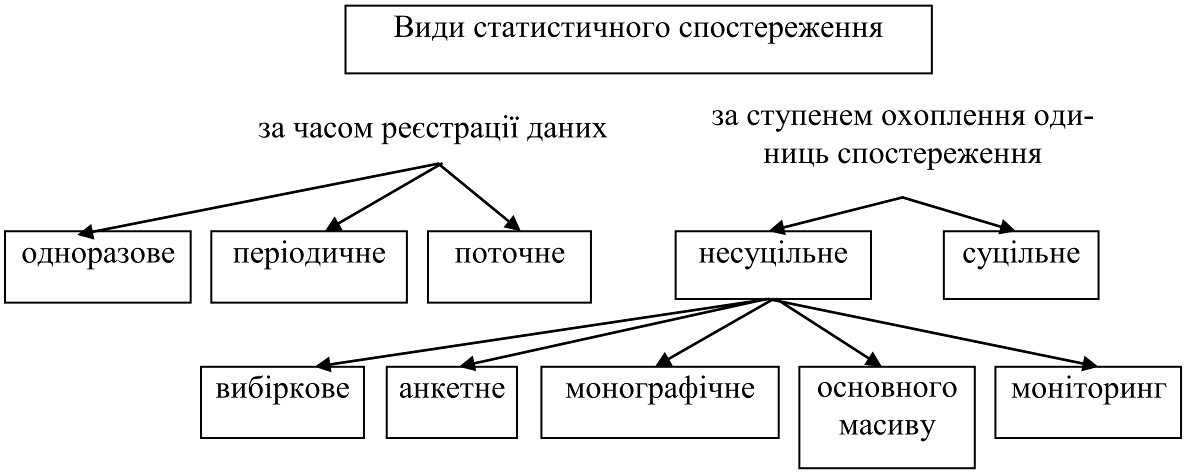

Рисунок 2.2

Способи спостереження

Статистичне спостереження здійснюється такими *трьома способами:* безпосередній облік фактів; документальний облік; опитування.

При *безпосередньому обліку фактів* відомості, що підлягають фіксації, певним чином підраховуються, виміряються, зважуються для одиниць об'єкта спостереження, наприклад: реєстрація товарних потоків, що перетинають кордон; облік готівкової грошової маси в банках тощо.

Документальне спостереження ґрунтується на використанні різних документів (звітності, бухгалтерських документів, річних звітів та ін.), якими визначаються показники на макро- та мікрорівні: обсяги матеріальних, трудових і фінансових ресурсів; розмір доходів; обсяги експорту та імпорту товарів та ін.

Опитування – такий спосіб спостереження, при якому відомості отримують усно або письмово зі слів опитуваних осіб. Опитування може бути експедиційним, кореспондентським та у формі самореєстрації.

За *експедиційним способом* реєстрація фактів здійснюється спеціально підготовленими обліковцями з одночасною перевіркою точності реєстрації (як, наприклад, під час перепису населення).

При *кореспондентському способі* спостереження потрібні відомості надають особи, які добровільно виявили бажання відповісти на поставлені в анкетах запитання; кореспондентський спосіб спостереження застосовується, наприклад, для дослідження ринку товарів і послуг окремих регіонів, для обстежування процесу просування товарів в умовах ринку тощо.

Самореєстрація – це реєстрація фактів самими респондентами після попереднього інструктажу з боку реєстраторів-обліковців; прикладом такого спостереження може бути бюджетне обстежування родин різних верств населення, при якому родини самі ведуть записи про свої доходи та витрати, а реєстратори-обліковці регулярно (двічі на місяць) відвідують їх, перевіряють повноту і правильність цих записів.

2.3. Помилки спостереження та методи їх контролю

У процесі збирання статистичного матеріалу можуть виникнути неточності, які називають *помилками спостереження.* Кількісно вони визначаються розбіжністю між дійсними розмірами ознак явищ і їх величиною, встановленою при спостереженні.

Розрізняють дві групи помилок статистичного спостереження: помилки реєстрації і помилки репрезентативності. Кожна з цих груп помилок поділяється на випадкові та систематичні.

Помилки реєстрації виникають внаслідок неправильного встановлення фактів у процесі спостереження або помилкового запису їх в формулярі.

Помилки репрезентативності виникають при вибірковому спостереженні через несуцільність реєстрації даних і порушення принципу випадковості відбору.

Випадкові помилки реєстрації пояснюються дією різних випадкових причин (описки, обмови, неточний підрахунок тощо). Ці помилки мають різну спрямованість і внаслідок закону великих чисел взаємно погашаються.

Систематичні помилки реєстрації виникають через дію певних постійних причин (свідоме перекручування фактів у бік зменшення або перебільшення їх величин, неточність вимірювальних приладів тощо). Такі помилки спрямовані в один бік і тому змінюють значення реєстрованих ознак.

Запобігти помилок спостереження можна за такими напрямами:

- використання наукового підходу до визначення об'єкта спостереження;
- ретельна розробка програми та організаційного плану спостереження;
- використання єдиної методології організації обліку і звітності;
- систематична перевірка органами статистики стану обліку і звітності на об'єктах;
- ретельний інструктаж обліковців і реєстраторів при проведенні переписів населення.

Контроль даних

Статистичний матеріал перевіряється з точки зору його повноти і правильності оформлення. Потім він підлягає контролю двох видів: логічного та арифметичного.

Суть *логічного контролю* полягає в перевіренні добутих даних між собою або з нормативними показниками. (Прикладом логічного контролю може бути порівняння відповідей респондентів про їх вік, сімейний стан, вид діяльності та джерела засобів існування.)

Арифметичний контроль полягає в арифметичній перевірці підсумкових та розрахункових показників, а також в арифметичній ув'язці пов'язаних між собою даних. (Наприклад, розмір акціонерного капіталу товариства можна визначити, коли відомі кількість акціонерів і розмір їхнього середнього внеску)

Контрольні запитання

1. Поняття статистичного спостереження.
2. Вимоги до статистичного спостереження.
3. Форми статистичного спостереження.
4. Види статистичного спостереження.
5. Способи статистичного спостереження.
6. Помилки статистичного спостереження.
7. Логічний та арифметичний контроль даних.

3. ЗВЕДЕННЯ, КЛАСИФІКАЦІЯ ТА ГРУПУВАННЯ СТАТИСТИЧНИХ ДАНИХ

3.1. Суть та організація статистичного зведення

Зведення як другий етап статистичного дослідження – це наукова обробка даних спостереження. *Основне завдання зведення* полягає у виявленні типових рис та закономірностей у сукупності.

Статистичні зведення розрізняють за рядом ознак: за складністю побудови; організацією роботи; ступенем обробки даних.

За *складністю побудови* зведення буває просте і складне. *Просте зведення* – це підбиття підсумків первинного матеріалу в цілому без будь-якої його систематизації. *Складне зведення* поєднує комплекс операцій: групування одиниць; підбиття групових і загальних підсумків; подання результатів зведення у формі статистичних таблиць, графіків, рядів розподілу.

За *організацією роботи* визначають централізоване і децентралізоване зведення. При *централізованому зведенні* всі дані зосереджуються в одному місці (центрі), де й обробляються (переписи населення, одноразові статистичні обстеження, соціологічні опитування). При *децентралізованому зведенні* обробка статистичної інформації здійснюється від нижчої до вищої ланки управління: від окремого підприємств до Держкомстату.

За *ступенем автоматизації обробки даних* розрізняють *автоматизоване* і *ручне* зведення.

3.2. Групування статистичних даних

Групування – розподіл сукупності масових явищ і процесів суспільного життя на типи і групи за найбільш характерними ознаками, які називають *групувальними*.

Групування, які спрямовані на виявлення зв'язку між окремими ознаками досліджуваного явища, називаються *аналітичними* (вивчаються взаємозв'язки між собівартістю та її факторами, продуктивністю праці та її факторами і ін.).

За кількістю групувальних ознак розрізняють прості та комбінаційні групування. *Простим* називають групування, яке проводиться за однією ознакою. У випадку поєднання двох і більше ознак групування є *комбінаційним.*

Якщо розмежування елементів сукупності на групи здійснюється за *атрибутивними ознаками,* то такий вид групування називають *класифікацією або номенклатурою.* Вони розробляються міжнародними та національними статистичними органами і рекомендуються як статистичний стандарт.

Прикладами діючих класифікацій національного рівня є такі, що повністю узгоджені з міжнародними стандартами:

- „Класифікація видів економічної діяльності” (КВЕД), де як ознака класифікації приймається одна з трьох ознак: призначення виробленої продукції; єдність технології виробництва; однорідність використаної сировини;
- „Українська класифікація товарів зовнішньоекономічної діяльності” (УКТ ЗЕД);
- „Державний класифікатор продукції та послуг” (ДКПП).

Якщо ознаками виступають *кількісні показники,* то такий вид робіт (на відміну від класифікацій) називають у вузькому розумінні безпосередньо *статистичним групуванням.*

При використанні методу групування вирішують такі питання:

- вибір групувальної ознаки;
- визначення кількості груп та величини інтервалу;
- встановлення переліку показників, якими повинні характеризуватися виділені групи стосовно конкретного групування;
- складання макетів таблиць, де будуть представлені результати групування;
- обчислення абсолютних, відносних і середніх показників;
- табличне і графічне оформлення результатів групування.

За групувальні приймають найістотніші ознаки. Групувальною ознакою може бути атрибутивна (якісна) або кількісна ознака. Наприклад, у табл. 3.1 ілюструється розподіл міських земель за видами забудови (атрибутивними ознаками).

Таблиця 3.1 – Розподіл земель міста за видами забудови

Вид міської забудови	Площа, га	% від підсумку
Промислова площа	670	22,1
Житлова	1920	63,5
Громадська	315	10,4
Інша	120	4,0
Всього	3025	100

Перші три групи таблиці представляють основні види забудови, а четверта – об'єднує решту менш вагомих видів.

Групування за атрибутивною ознакою називають також *атрибутивним рядом розподілу* (табл. 3.2); за кількісними – *варіаційним рядом розподілу*.

Елементами (характеристиками) атрибутивного/варіаційного *ряду* розподілу є: *значення атрибутивної/варіаційної ознаки*; *частоти* f_i – числа, які показують, як часто зустрічається те чи інше значення ознаки в ряду, *частки* φ_i – це частоти, виражені у відносних величинах (коефіцієнтах або процентах).

Варіаційні ряди розподілу (далі – *варіаційні ряди*) бувають дискретними та інтервальними.

Дискретні варіаційні ряди основані на величинах ознак, що мають цілі значення (наприклад, тарифний розряд робітників, розподіл сімей за кількістю дітей тощо, див. табл. 3.3).

В інтервальних варіаційних рядах групувальна ознака може приймати будь-яке значення (ціле, дробове) в межах кожного інтервалу (наприклад, розподіл заробітної плати працюючих).

Крім таких елементів варіаційних рядів як частота і частка, використовуються показники *щільність частоти* (визначається співвідношенням f_i/h, де h – ширина інтервалу) та *накопичена частота* ($F=\sum_{i=1}^{k}\varphi_i$, де k – порядковий номер інтервалу) (табл. 3.4).

Таблиця 3.2 – Розподіл студентів університету за економічними спеціальностями

Назва спеціальності	Чисельність студентів, осіб, f_i	У % від загальної кількості, φ_i
Фінанси	262	27,3
Облік і аудит	279	29,0
Менеджмент організацій	246	25,6
Маркетинг	174	18,1
Всього	961	100

Таблиця 3.3 – Розподіл сімей в населеному пункті за кількістю дітей на 01.12.2002 року

Кількість дітей, x_i	Кількість сімей, f_i	У % від загальної кількості сімей, φ_i
1	63	56,2
2	48	42,8
3 і більше	1	1,0
Всього	112	100

Таблиця 3.4 – Розподіл господарств за поголів'ям крупної рогатої худоби за звітний період

Кількість голів, x_i	Кількість господарств, f_i	Частка господарств, φ_i, %	Щільність розподілу, f_i/h	Накопичена частка, F
До 300	100	50	0,334	50
300–599	40	20	0,134	70
600–999	30	15	0,100	85
1000–3000	10	5	0,005	90
Більше 3000	20	10	–	100
Всього	200	100	X	X

3.3. Інтервали групування

При групуванні на основі *кількісних ознак* (дискретних або неперервних) визначають кількість груп та інтервали групування.

Для визначення *кількості груп* необхідно дотримуватися двох важливих умов побудови групувань:

1) виділені групи мають відрізнятися якісною однорідністю;

2) кількість одиниць у кожній групі має бути досить великою, що відповідає вимозі закону великих чисел.

У масових сукупностях оптимальну кількість груп з рівними інтервалами приблизно можна визначити за формулою американського вченого Стерджеса

$$m = 1 + 3{,}322 \lg n, \tag{3.1}$$

де m – кількість інтервалів; n – обсяг сукупності.

Формула Стерджеса може бути використана при умові, що розподіл одиниць сукупності за даною ознакою наближається до нормального закону розподілення.

Інтервали, тобто проміжок між крайніми значеннями ознаки в групі одиниць, бувають *рівні, нерівні, відкриті* та *закриті*. Вибір виду інтервалу залежить від характеру розподілу одиниць досліджуваної сукупності.

Рівні інтервали використовують у випадках, коли значення варіюючої ознаки x змінюються плавно, поступово, рівномірно. Ширина інтервалу h визначається за формулою

$$h = \frac{x_{\max} - x_{\min}}{m}, \tag{3.2}$$

де $x_{\min}$, $x_{\max}$ – найменше та найбільше значення ознаки у сукупності.

Наприклад, прибутковість активів комерційних банків коливається в межах від 5 до 45 %. При прийнятій кількості груп m = 4 ширина інтервалу h = 4. Тоді межі інтервалів становлять відповідно:
5–15, 15–25, 25–35, 35–45. Наведений розподіл прибутковості активів банків може бути представлено в іншому вигляді: до 15, 15–25, 25–35, 35 і

більше. Перший та останній інтервали мають лише одну межу і називаються *відкритими*, інші – *закритими*.

Нерівні інтервали використовуються тоді, коли діапазон значень ознаки надто широкий і розподіл сукупності за цією ознакою нерівномірний. Наприклад, розподіл селищ міського типу за кількістю жителів (тис. чол.): до 3; 3–4,9; 5–9,9; 20–49,9.

3.4. Статистичні таблиці

Результати статистичного зведення та групування, як правило, оформлюють у вигляді статистичних таблиць. *Статистична таблиця* – це форма представлення числових даних, які характеризують досліджувані явища і процеси.

Статистичні таблиці поділяють на прості, групові та комбінаційні.

У простій таблиці кожній одиниці явища, що формується за однією ознакою, відповідає одна група статистичних даних (табл.3.5).

Таблиця 3.5 – Наявність будівельних машин у будівельних управліннях регіону

Вид машин	Кількість машин даного виду, тис. шт.
Екскаватори	3,6
Скрепери	0,7
Бульдозери	3,1
Крани пересувні	4,3
Всього	11,7

У групових таблицях кожній одиниці явища, що формується за однією ознакою, відповідає декілька груп статистичних даних (табл. 3.6). У комбінаційних таблицях кожній одиниці явища, що формується за декількома ознаками, відповідає декілька груп статистичних даних (табл. 3.7).

Складання статистичної таблиці здійснюється у два етапи. На першому етапі розробляється макет таблиці, на другому – заповнюється статистичними даними.

Таблиця 3.6 – Групування магазинів за рівнем продуктивності праці працівників за звітний період

Рівень продуктивності роботи магазинів, тис. грн	Кількість магазинів	Фондовіддача на 1 грн активної частки основних фондів, грн	Рентабельність активної частки основних фондів, %
До 60	4	40,42	2,3
60–70	4	43,1	2,8
70–80	7	75,8	4,7
80–90	7	65,9	4,0
90–100	3	93,1	5
Більше 100	7	109,3	6,4
Всього	32	**Х**	**Х**
У середньому	**Х**	75	4,4

Таблиця 3.7 – Групування продовольчих магазинів міста за часткою площі торгівельного залу та тривалістю робочого дня

Групи та підгрупи магазинів за часткою площі торгівельного залу (%) та тривалістю робочого дня (год)	Кількість магазинів	Фондовіддача на 1 грн активної частки основних фондів, грн	Рентабельність активної частки основних фондів, %
До 35 %	13	48,5	3,10
У тому числі: 8–10 год. більше 10 год	4 9	41,2 57,5	2,20 4,02
35–45%	21	69,8	5,20
У тому числі: 8–10 год. більше 10 год.	6 15	54,6 77,4	3,08 7,10
45–55 %	18	90,6	6,40
У тому числі: 8–10 год. більше 10 год.	5 13	68,9 108,7	4,17 7,98
Всього	52	**Х**	**Х**
У середньому	**Х**	73,5	4,70

3.5. Статистичні графіки

Статистичний графік – це зображення статистичних даних за допомогою ліній, геометричних фігур та інших наочних засобів.

Статистичні графіки можна поділити на три групи: *діаграми, статистичні карти, ряди розподілу*.

Найбільш поширеною групою є *діаграми,* на яких статистичні дані зображуються за допомогою геометричних знаків, ліній і фігур. Основні види діаграм: лінійні, радіальні, секторні, стовпчикові, стрічкові, фігурні та інші.

Динаміку (розвиток явищ у часі) найчастіше відображають за допомогою *лінійних діаграм*: на осі абсцис відкладають періоди або моменти часу, на осі ординат – числові значення показника.

Секторні діаграми характеризують структуру явища. Для побудови секторної діаграми круг розділяється радіусами на сектори, площі яких пропорційні частинам досліджуваного явища в загальному обсязі зображуваного круга.

Радіальні діаграми використовують для зображення явищ, які періодично змінюються за часом (переважно сезонних коливань). Для їх побудови застосовують полярну систему координат.

Вертикальні стовпчики в масштабі відповідають чисельним значенням ознаки. *Стовпчики* можуть розташовуватися один від одного на однаковій відстані або щільно. Зображення стовпчиків може бути площинним або об'ємним. Якщо стовпчики розташовуються не вертикально, а горизонтально, то такі діаграми називаються *стрічковими (смугастими).* У статистиці, перш за все в рекламних цілях, використовують також *фігурні діаграми.* При їх побудові статистичні дані зображуються малюнками-символами (банки консервів, автомобілі тощо), площі яких пропорційні величинам відповідних ознак. Ці діаграми більш наочні, легше сприймаються і тому їх використовують для реклами окремих товарів.

Другою групою статистичних графіків є *статистичні карти – картограми* і *картодіаграми*.

Картограми – це зображення певної ознаки на схематичній географічній карті різними забарвленням або штрихуванням. Наприклад, різна щільність населення країни може бути відображена різною інтенсивністю забарвлення території.

Картодіаграма являє собою поєднання схематичної географічної карти з однією із згаданих вище діаграм. При цьому статистичні показники зображуються у вигляді стовпчиків, квадратів, трикутників, силуетів тощо.

Особливе місце, у зв'язку із специфічністю, займає *графічне зображення рядів розподілу.*

Для графічного зображення дискретного варіаційного ряду використовують *полігон розподілу* (рис. 3.1). Його зображують у прямокутній системі координат, де на осі абсцис відкладають значення варіант x, а на осі ординат – частоти f. Одержані точки з координатами x_i та f_i з'єднують прямими лініями. Для замикання полігону кінцеві вершини з'єднують з точками на осі абсцис (див. лінії на рис. 3.1), які віддалені на одну поділку від x_{max} і x_{min}.

Графічне зображення інтервального варіаційного ряду виконують у вигляді *гістограми.* Для рядів з рівними інтервалами будують гістограму в осях „$x - f$" (рис. 3.2). Для незакритого першого інтервалу у якості x_1 беруть середнє значення другого інтервалу, а для незакритого останнього інтервалу – середнє значення передостаннього інтервалу x_{n-1}.

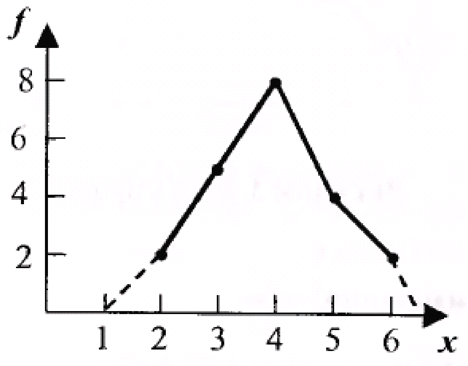

Рисунок 3.1

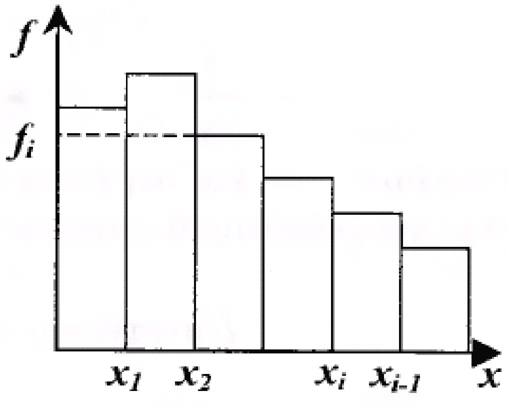

Рисунок 3.2

Контрольні запитання

1. Поняття статистичного зведення та його види.
2. Поняття статистичного групування та його види.
3. Поняття інтервалу групування та його види.
4. Поняття ряду розподілу та його елементи.
5. Класифікація варіаційних рядів.
6. Поняття статистичної таблиці та її види.
7. Поняття статистичного графіка.
8. Діаграми та їх види.

4. СТАТИСТИЧНІ ПОКАЗНИКИ

4.1. Види, типи та значення статистичних показників

Після зведення та групування даних спостереження переходять до останнього – *третього етапу* статистичної методології. Він полягає в подальшій обробці статистичних таблиць шляхом обчислення статистичних показників.

Статистичний показник – це узагальнююча характеристика явища або процесу, яка характеризує всю сукупність одиниць обстеження і використовується для аналізу сукупності в цілому. За допомогою статистичних показників вирішується одна з головних задач статистики: визначається кількісна сторона явища чи процесу у поєднанні з якісною стороною.

Кількісна сторона показника подається числом з відповідною одиницею вимірювання для характеристики: розміру явищ (кількості робітників, обсягу товарообороту, капіталу фірми тощо); їх рівнів (наприклад, рівня продуктивності праці робітників); співвідношень (наприклад, між продавцями та іншими категоріями працівників магазину).

Якісний зміст показника залежить від суті досліджуваного явища (процесу) і відображається у назві показника (прибутковість, народжуваність тощо).

Показники поділяються на *види* залежно від способу їх обчислення, ознак часу, виконання своїх функцій.

За *способом обчислення* розрізняють первинні та похідні показники. *Первинні* визначаються шляхом зведення та групування даних і подаються у формі абсолютних величин (наприклад, кількість та сума вкладів громадян у банку). *Похідні* показники обчислюються на базі первинних і мають форму середніх або відносних величин (наприклад, середня заробітна плата, індекс цін).

Серед статистичних показників окрему групу становлять *взаємообернені показники* – пара характеристик, які існують паралельно і відповідають одному і тому ж явищу (процесу). Це *прямий* показник *х,* який змінюється у напрямку зміни явища (наприклад, продуктивність праці за одну одиницю часу), та обернений $1/x$ – у протилежному напрямку (наприклад, трудомісткість одиниці продукції).

За *ознакою часу* показники поділяються на інтервальні та моментні. *Інтервальні* характеризують явище за певний період часу (місяць, квартал, рік): наприклад, середньомісячні сукупні витрати на душу населення. *Моментні* показники характеризують явище за станом на певний момент часу (дату): наприклад, залишок обігових коштів на початок місяця.

За способом виконання своїх функцій розглядають показники, що відбивають обсяг явища, його середній рівень, інтенсивність прояву, структуру, зміну в часі або порівнянні у просторі.

У статистиці використовують декілька різновидів статистичних показників:

- абсолютні та відносні величини;
- середні величини;
- показники варіації.

4.2. Абсолютні та відносні величини

Абсолютними величинами називають кількісні показники, які визначають рівень, обсяг, чисельність суспільних явищ (наприклад, капітал фірми на початок року, посівна площа сільськогосподарських підприємств на даний момент часу, чисельність робітників підприємства у звітному періоді тощо).

За способом вираження явища, що вивчається, абсолютні величини розподіляються на індивідуальні та загальні (сумарні). *Індивідуальні величини* характеризують ознаки окремих одиниць сукупності (наприклад, розмір заробітної плати окремого робітника, кількість заявок та обсяги попиту на купівлю товару товарної біржі та ін.) *Загальними величинами* є такі абсолютні показники, які знаходять при сумуванні індивідуальних абсолютних величин (наприклад, фонд заробітної плати робітників підприємств району, вартість основних фондів сільськогосподарських підприємств області тощо).

Абсолютні величини – це іменовані числа і залежно від характеру явища або процесу можуть мати різні *одиниці* вимірювання: *натуральні* (кг, м, шт. і т.д.); *умовно-натуральні* (одна умовна банка консервів, одна умовна одиниця мінеральних добрив і т.д.); *трудові* (людино-година, людино-день); *вартісні* (грн, руб., дол. США, євро та ін.).

Абсолютні показники відіграють важливу роль у статистиці, однак вони не можуть дати достатньо повного уявлення про досліджуване явище. Тому виникає потреба в обчисленні інших узагальнюючих показників – відносних та середніх величин, підґрунтям для яких є абсолютні величини.

Відносні величини – це узагальнюючі кількісні показники, які виражають співвідношення порівнюваних абсолютних величин.

Формулою відносної величини є такий звичайний дріб:

$$\text{Відносна величина} = \frac{\text{Величина порівняння}}{\text{База порівняння}}.$$

Залежно від величин чисельника та знаменника цього дробу відносні величини можуть бути виражені у таких формах: *коефіцієнтах* (частках), *процентах* (%), *проміле* (‰), *продециміле* ($^{0}/_{000}$), коли за базу порівняння приймають відповідно 1, 100, 1000, 10000 одиниць.

Різноманітність співвідношень у реальному житті потребує різних за змістом і статистичною природою відносних величин. Залежно від своїх функцій відносні величини можна класифікувати так:

1) відносні величини динаміки;
2) відносні величини структури;
3) відносні величини координації;
4) відносний показник планового завдання;
5) відносний показник виконання плану;
6) відносні величини інтенсивності;

1) *Відносна величина динаміки*

Динамікою у статистиці називають зміну явища в часі. *Відносна величина динаміки* характеризує напрям та інтенсивність зміни показника за часом і визначається співвідношенням його значень за два періоди або моменти часу. При цьому базою порівняння може бути змінний попередній рівень (розрахунок *ланцюговим способом)* або постійний віддалений за часом рівень (розрахунок *базисним способом).* Відносні показники динаміки називають *темпами зростання.* Наприклад, розмір інвестицій у галузь становив у млн. грн: 2002 р. – 420,0; 2003 р. – 546,0; 2004 р. – 573,5. Порівнюючи значення показника, отримаємо темпи зростання інвестицій:

- *розрахунок ланцюговим способом:* у 2003 р. порівняно з 2002 р. 546,0/420,0 =1,30 або 130 % (інвестиції зросли на 30 %); у 2004 р. порівняно з 2003р. 573,5/546,0 = 1,05 або 105 % (інвестиції зросли на 5 %);

- *розрахунок базисним способом:* якщо за базу приймається рівень інвестицій у 2002 р., то у 2003 р. темп зростання буде 1,3 або 130 %; у 2004 р. порівняно з базовим рівнем у 2002 р. темп зростання 573,5/420,0 = 1,365 або 136,5 % (інвестиції зросли на 36,5 %).

Якщо значення показника зменшується, то величина динаміки буде меншою за одиницю.

2) *Відносна величина структури*

Статистичні сукупності завжди структуровані і мають певні складові. *Відносна величина структури* характеризує склад, структуру сукупності за тією чи іншою ознакою і показує внесок складових сукупності до загальної маси. Вона визначається відношенням розмірів складових частин сукупності до загального підсумку. Скільки складових, стільки відносних величин структури. Вони визначаються простим чи десятковим дробом або процентом. Наприклад, частка осіб допрацездатного віку міста становить 0,25, або 25 %.

3) *Відносна величина координації*

Відносна величина координації дає співвідношення різних структурних одиниць тієї самої сукупності і показує, скільки одиниць однієї частини сукупності припадає на 1, 100, 1000 і більше одиниць іншої, взятої за базу порівняння. Наприклад, частка власних коштів фірми становить 70 %, а залучених – 30 %. Тоді відносна величина координації може складати 30/70 = 0,43, а це означає, що на одиницю власних коштів припадає 0,43 залучених. Інший приклад: скільки чоловіків припадає на 1000 жінок.

4) *Відносні показники планового завдання та* 5) *виконання плану*

Відносний показник планового завдання – це відношення величини показника, встановленого на плановий період, до його величини, досягнутого за попередній період, який взято за порівняльну базу. Наприклад, на сільськогосподарському підприємстві середньорічний надій від корови у плановому періоді встановлено 3320 кг, за попередній (базисний) рік було 3200 кг. Тоді відносний показник планового завдання дорівнює: $K_{\text{пз}}$ =3320/3200=1,038, тобто у плановому періоді надій молока очікується на 3,8 % більше, ніж у базисному періоді.

Відносний показник виконання плану являє собою відношення фактично досягнутого рівня до планового завдання. Наприклад, у періоді, що планується, середньорічний надій молока від корови фактично становив 3480 кг. У такому випадку, звертаючись до попереднього прикладу, відносний показник виконання плану становить: $K_{\text{вп}}$ =3480/3320 = 1,048, тобто, фактично у розглянутому періоді надій молока на 4,8 % більше плану.

Відносні показники динаміки (К), планового завдання ($K_{\text{пз}}$*)* та *виконання плану (*$K_{\text{вп}}$*)* пов'язані між собою такими рівняннями: $K = K_{\text{пз}} \cdot K_{\text{вп}}$. За нашими прикладами $K = 1{,}038 \cdot 1{,}048 = 1{,}088$, або K = 3480/3200 = 1,088.

6) *Відносна величина інтенсивності*

Відносна величина інтенсивності характеризує відношення різноіменних величин, по'язаних між собою певним чином. Це – щільність населення на 1 км^2 (наприклад, 82,5 осіб/ км^2), виробництво електроенергії на душу населення (наприклад, $5625 \text{ кВт} \cdot \text{год/осіб}$) тощо. Якщо обсяги явища незначні відносно обсягів середовища, то їх співвідношення збільшуються у 100, 1000 і більше разів. Наприклад, показники народжуваності, смертності, шлюбності розраховуються на 1000 осіб населення, захворюваність та злочинність – на 100000 осіб населення.

4.3. Середні величини

Середньою величиною в статистиці називаються кількісний показник характерного, типового рівня масових однорідних явищ. У зв'язку з цим середні величини належать до узагальнюючих статистичних показників, які дають зведену, підсумкову характеристику масових суспільних явищ. У середній величині гасяться (розчиняються) всі відмінності та особливості індивідуальних значень ознак і вона є „рівнодіючою" значень цих ознак.

Головними умовами застосування середніх величин є:

1) наявність *якісної однорідності* сукупності;

2) *масовий характер даних* сукупності, де діє закон великих чисел.

Залежно від характеру ознаки, що усереднюється, і наявності вихідної статистичної інформації в статистиці використовують декілька видів середніх, серед яких найбільш поширеними є такі: *середня арифметична; середня гармонічна; середня квадратична; середня геометрична.* Поряд з переліченими видами середніх величин у статистичній практиці застосовують також *середню хронологічну* (її обчислення розглянемо пізніше) та структурні середні: *моду та медіану.*

Кожна із зазначених видів середніх може виступати у двох *формах:* простої та зваженої. *Проста середня* застосовується при обчисленні середньої за первинними (не згрупованими) даними, *зважена* – за згрупованими даними.

При використанні середніх величин введемо такі позначення:
$\bar{x}$ – середнє значення досліджуваної ознаки; x_i або x – кожне індивідуальне значення усереднюваної ознаки (варіанта) у варіаційному ряду; f_i або f – частота повторень (вага) індивідуальної ознаки у варіаційному ряду; $z=xf$ – обсяг значень ознаки; n – кількість одиниць досліджуваної ознаки.

Середня арифметична

Середня арифметична - це найпоширеніший вид середньої між інших. Вона застосовується тоді, коли відомі індивідуальні значення усереднюваної ознаки та їх кількість у сукупності. Тоді *проста середня арифметична* обчислюється діленням загального обсягу значень ознаки на обсяг сукупності (далі замість $\sum_{i=1}^{n} x_i$ буде $\sum x$)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{4.1}$$

Наприклад, статутний капітал акціонерної компанії сформований 6 засновниками. Розмір внеску кожного з них відповідно становив, млн грн: 8; 10; 12; 9; 6; 5. Середній внесок одного засновника розраховується так.

Зважена середня арифметична використовується у тих випадках, коли значення ознаки подано у вигляді варіаційного ряду, в якому чисельність одиниць у варіантах неоднакова. Формула середньої арифметичної зваженої має вигляд

$$\bar{x} = \frac{\sum fx}{\sum f}. \tag{4.2}$$

Із формули (4.2) видно, що середня зважена принципово не відрізняється від середньої простої арифметичної (4.1). Тут додавання f разів величини x змінюється множенням її на кількість повторень (f).

Техніку обчислення середньої арифметичної зваженої проілюструємо прикладом обчислення середньої виробки деталей на одного робітника за зміну, якщо відомо скільки деталей виготовив кожен з 15 робітників (табл. 4.1):

Таблиця 4.1 – Розподіл робітників за виробкою деталей

Виробка деталей за зміну одним робітником, шт. x	Кількість робітників (ваги), f	xf
18	2	36
19	4	76
20	5	100
21	3	63
22	1	22
Всього	15	297

За формулою (4.2) середня арифметична зважена:

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{297}{15} = 19{,}8 \approx 20 \text{ шт.}$$

Середня гармонічна

Середня гармонічна – це обернена до середньої арифметичної із обернених значень ознак. Її обчислюють, коли необхідне осереднення обернених індивідуальних значень ознак шляхом їх підсумування (наприклад, у випадках визначення середніх витрат часу, матеріалів на одиницю продукції тощо). У випадку розрахунку середньої гармонічної зваженої її обчислюють тоді, коли відомі дані про загальний обсяг ознаки ($z=xf$), а також індивідуальні значення ознаки (x), невідомою є частота (f). Формули середньої гармонічної – простої і зваженої – мають такий вигляд:

$$\text{для простої } \bar{x}_h = \frac{n}{\sum \frac{1}{x}}; \quad \text{зваженої } \bar{x}_h = \frac{z}{\sum \frac{z}{x}}. \qquad (4.3)$$

Для встановлення місця середньої гармонічної в розрахунку середньої величини розглянемо такий приклад. Припустимо, що бригада токарів протягом 8 – годинного робочого дня зайнята обточкою однакових деталей. Перший токар затрачує на одну деталь 12 хв, другий – 15 хв, третій – 11 хв, четвертий – 16 хв і п'ятий – 14 хв. Необхідно знайти середній час на виготовлення одної деталі.

На перший погляд, ця задача розв'яуються легко за формулою середньої арифметичної простої

$$\bar{x} = 13{,}6 \text{ хв.}$$

Однак, знайдена середня була би правильною, якщо кожний робітник виробив тільки по одній деталі, а не працював 8 годин, коли робітниками була виготовлена різна кількість деталей. Для розрахунку кількості деталей, виготовлених кожним робітником, використаємо таке співвідношення (логічну формулу):

$$\text{Середній час на одну деталь} = \frac{\text{Весь витрачений час}}{\text{Кількість деталей}} =$$

$$= \frac{8\cdot 60 + 8\cdot 60 + 8\cdot 60 + 8\cdot 60 + 8\cdot 60}{\frac{8\cdot 60}{12} + \frac{8\cdot 60}{15} + \frac{8\cdot 60}{11} + \frac{8\cdot 60}{16} + \frac{8\cdot 60}{14}} = \frac{5}{\frac{1}{12} + \frac{1}{15} + \frac{1}{11} + \frac{1}{16} + \frac{1}{14}} = 13{,}3 \text{ хв.}$$

Останнє кількісне співвідношення відповідає формулі середньої гармонічної $\bar{x}_h$ простої.

Бачимо, що в наявності різниця між результатами обчислення за формулами середньої арифметичної та середньої гармонічної.

Середня квадратична

Середня квадратична використовується для визначення показників варіації (коливання) ознаки – дисперсії та середнього квадратичного відхилення. Обчислюється на основі квадратів відхилень індивідуальних значень ознаки від їх середньої величини. Формула середньої квадратичної має такий вигляд:

$$\text{проста } \bar{x} = \sqrt{\frac{\sum x^2}{n}}\text{; зважена } \bar{x} = \sqrt{\frac{\sum x^2 f}{\sum f}}. \qquad (4.4)$$

Середня геометрична

Середню геометричну застосовують у тих випадках, коли обсяг сукупності формується не сумою, а добутком індивідуальних значень ознак. Цей вид середньої використовується здебільшого для обчислення середніх коефіцієнтів (темпів) зростання в рядах динаміки. Так, у випадку однакових часових інтервалів між рівнями динамічного ряду середня геометрична проста має такий вигляд:

$$k = \sqrt[n]{k_1 k_2 \ldots k_n}\,, \tag{4.5}$$

де $k_i = y_i / y_{i-1}$ – темпи зростання; y_{i-1} – попередній до y_i-го рівень ряду; n – кількість інтервалів.

Прикладом застосування середньої геометричної є такий. Припустимо, що внаслідок інфляції споживчі ціни за чотири роки зросли в 2,8 раза, в тому числі: за перший рік у 1,7 рази; за другий – в 1,3; за третій – в 1,1; за четвертий – в 1,15 рази. Як визначити середньорічний темп зростання цін? Середня арифметична (1,7+1,3+1,1+1,15):4=1,312 не забезпечує визначеної властивості, оскільки за чотири роки за цією середньою ціни зросли б у 1,312·1,312·1,312·1,312=2,94 раза, а не в 2,8 раза. Визначену властивість забезпечує тільки середня геометрична $\bar{x} = \sqrt[4]{1{,}7 \cdot 1{,}3 \cdot 1{,}1 \cdot 1{,}15} = 1{,}295.$

Мода і медіана

Середніми величинами в статистичних рядах розподілу є *мода і медіана,* які належать до класу *структурних (позиційних) середніх.* Їх величини залежать лише від характеру частот, тобто від структури розподілу. На відміну від інших середніх, які залежать від усіх значень ознаки, мода і медіана не залежать від крайніх значень. Це особливо важливо для незакритих крайніх інтервалів варіаційних рядів розподілу.

Мода (M_o) – це значення варіанти, що найчастіше повторюється в ряду розподілу. Спосіб обчислення моди залежить від виду статистичного ряду. Для *атрибутивних і дискретних рядів* розподілу моду визначають візуально без будь-яких розрахунків за значенням варіанти з найбільшою частотою (часткою). Наприклад, за результатами опитування населення щодо самовизначення матеріального стану за чотирма оцінками (добрий, задовільний, незадовільний, нестерпний) більшість респондентів визначили свій стан як незадовільний – це і буде модою. Або модальною ціною на той чи інший продукт на ринку є та ціна, яка спостерігається найчастіше. В інтервальному ряді спочатку визначається модальний інтервал (інтервал з найбільшою частотою) і значення моди в середині інтервалу розраховується за формулою

$$M_{\text{o}} = x_0 + h\frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)}, \tag{4.6}$$

де x_0 – нижня межа модального інтервалу; h – величина модального інтервалу; f_1, f_2, f_3 – частота відповідно передмодального, модального та післямодального інтервалів.

Медіаною (M_{e}) називають варіанту, що ділить ранжований (впорядкований за мірою зростання або зменшення) ряд на дві рівні за обсягом частини. Медіана для дискретного ряду з непарним числом варіант буде відповідати середній варіанті $M_{\text{e}} = x_{m\text{-}1}$, де m – номер кратної варіанти першої половини ранжованого ряду. Медіана для дискретного ряду з парним числом варіант буде відповідати середній із значень варіант у ранжованому ряду: $M_{\text{e}} = \frac{x_m - x_{m+1}}{2}$. Для інтервального ряду медіана обчислюється для середини медіанного інтервалу, за який приймається такий, де сума накопичених частот перевищує половину значень частот ряду розподілу. В даному випадку формула для розрахунку медіани має вигляд

$$M_{\text{e}} = x_0 + h\frac{0{,}5\sum f - S_{Me-1}}{f_m}, \tag{4.7}$$

де x_0 – нижня межа медіанного інтервалу; h – величина медіанного інтервалу; $0{,}5\sum f$ – половина суми накопичених частот інтервального ряду; S_{Me-1} – сума накопичених частот перед медіанним інтервалом; f_m – частота медіанного інтервалу.

Для аналізу закономірностей розподілу використовуються також такі характеристики як квартилі та децилі. *Квартилі* – це варіанти, які поділяють обсяги сукупності на чотири рівні частини, *децилі* – на десять частин.

4.4. Показники варіації

Після встановлення середньої величини *(*$\bar{x}$*, M_o, M_e)* виникає питання, в якій мірі індивідуальні значення ознаки відрізняються між собою та від середньої. Для цього розраховують показники варіації.

Варіацією ознаки називають різницю у числових значеннях ознак одиниць сукупності та їх коливання навколо середньої величини, що характеризує сукупність. Чим менша варіація, тим однорíднішою є сукупність і більш надійною (типовою) є середня величина.

До основних абсолютних і відносних показників, що характеризують варіацію, належать такі: розмах варіації; середнє лінійне відхилення; дисперсія; середнє квадратичне відхилення; коефіцієнт варіації тощо.

Розмах варіації – це різниця між найбільшим та найменшим значеннями ознаки

$$R = x_{\max} - x_{\min}. \tag{4.8}$$

Величина показника залежить тільки від крайніх значень ознаки і не враховує всіх значень, що знаходяться між ними.

Досконалішим є визначення варіації через інші показники, які дають змогу усунути недолік розмаху варіації.

Середнє лінійне відхилення являє собою арифметичну з абсолютних значень усіх відхилень індивідуальних значень ознаки від середньої

$$\text{просте } d = \frac{|x - \bar{x}|}{n}; \qquad \text{зважене } d = \frac{\sum f|x - \bar{x}|}{\sum f}. \tag{4.9}$$

Наявність абсолютних значень відхилень від середньої пояснюється так: середня арифметична має нульову властивість, відповідно якої сума відхилень індивідуальних значень ознаки зі своїми знаками дорівнює нулю; щоб мати суму всіх відхилень, відмінних від нуля, кожне з них слід брати за абсолютною величиною.

Основним недоліком середнього лінійного відхилення є те, що в ньому не враховуються знаки відхилень, тобто їх спрямованість. Тому цей показник варіації використовується рідко (аналіз складу працюючих, ритмічність виробництва, обіг коштів у зовнішній торгівлі тощо).

Показниками варіації, які усувають недоліки середнього лінійного відхилення, є дисперсія та лінійне квадратичне відхилення.

Дисперсією називають середню арифметичну квадратів відхилень індивідуальних значень ознаки. В залежності від вихідних даних дисперсія може обчислюватись за формулами середньої арифметичної простої або зваженої

$$\text{проста } \sigma^2 = \frac{\sum (x-\bar{x})^2}{n}; \text{ зважена } \sigma^2 = \frac{\sum f(x-\bar{x})^2}{\sum f}. \qquad (4.10)$$

Дисперсія – це один з найбільш розповсюджених в економічній практиці узагальнюючих показників розміру варіації у сукупності. Дисперсію використовують не лише для оцінки варіації, а й для вимірювання зв'язків між досліджуваними факторами; розклад дисперсії на складові дозволяє оцінити вплив різних факторів, які обумовлюють варіацію ознаки.

Середнє квадратичне відхилення, як і дисперсія, виступає як узагальнений показник варіації. Його обчислюють, отримавши квадратичний корінь з дисперсії: $\sigma = \sqrt{\sigma^2}$.

Смислове значення середнього квадратичного відхилення таке саме, як і лінійного відхилення: воно показує, на скільки в середньому відхиляються індивідуальні значення ознаки від їх середнього значення. Перевага цього показника порівняно із середнім лінійним відхиленням полягає у відсутності умовного припущення з сумування відхилень без врахування їх знаків, бо відхилення використовуються у квадратному степені. Крім зазначеного, перевагою даного показника у зіставленні з дисперсією є те, що середнє квадратичне відхилення виражається в тих же одиницях вимірювання, що і значення досліджуваної ознаки (грн, кг, га тощо). Тому цей показник називають також *стандартним відхиленням.*

У статистичній практиці часто виникає необхідність порівняння варіацій різних ознак. Наприклад, великий інтерес має порівняння віку робітників з їх кваліфікацією, стажу роботи з розміром заробітної плати, собівартістю та прибутку і ін. При таких порівняннях розглянуті показники коливання ознак з різними одиницями вимірювання не можуть бути використані (наприклад, неможливо порівнювати коливання стажу роботи в роках з варіацією заробітної плати в гривнях).

Для здійснення такого роду порівнянь, а також при зіставленні ознаки у декількох сукупностях з різними середніми арифметичними використовують відносний показник варіації – коефіцієнт варіації.

Коефіцієнтом варіації називають процентне відношення середнього квадратичного відхилення до середньої арифметичної величини ознаки

$$V_\sigma = \frac{\sigma}{\bar{x}} 100\,\%. \tag{4.11}$$

Чим більший коефіцієнт варіації, тим менш однорідна сукупність і тим менш типова середня для даної сукупності: сукупність вважають *кількісно однорідною,* якщо коефіцієнт варіації не перевищує 33 %.

Дисперсія посідає особливе місце у статичному аналізі соціально-економічних явищ і є важливим елементом статистичних методів, зокрема у *дисперсному аналізі.*

У структурованій сукупності, яка поділена на m груп за факторною ознакою x, загальна дисперсія σ^2 результативної ознаки y може бути представлена складовими: міжгрупова дисперсія δ^2 та середня з групових дисперсій $\bar{\sigma}^2$. Згідно з *правилом розкладання дисперсій* має місце рівняння

$$\sigma^2 = \delta^2 + \bar{\sigma}^2. \tag{4.12}$$

Загальна дисперсія σ^2 вимірює варіацію результативної ознаки в цілому за сукупністю під впливом усіх факторів, які обумовлюють цю варіацію. Загальна дисперсія для зваженої результативної ознаки y обчислюється за формулою (4.10).

Міжгрупова дисперсія δ^2 характеризує варіацію ознаки y за рахунок фактора x, покладеного в основу групування, і розраховується за формулою

$$\delta^2 = \frac{\sum_{j=1}^{m} (\bar{y}_j - \bar{y})^2 f_j}{\sum_{j=1}^{m} f_j}, \tag{4.13}$$

де $\bar{y}_j$, $\bar{y}$ – відповідно середня j-ї групи та загальна середня варіюючої ознаки; f_j – чисельність одиниць (частота) j-ї групи, m – кількість груп.

Для розрахунку середньої з групових дисперсій спочатку обчислюється *внутрішньогрупова дисперсія,* яка характеризує варіацію результативної ознаки за рахунок інших факторів, не врахованих у групуванні:

$$\sigma_j^2 = \frac{\sum_{j=1}^{m}\left(y_j - \bar{y}_j\right)^2}{f_j}. \qquad (4.14)$$

Для всіх груп у цілому розраховується середня з *групових дисперсій,* зважених на частоти відповідних груп:

$$\bar{\sigma}^2 = \frac{\sum_{j=1}^{m}\sigma_j^2 f_j}{\sum_{j=1}^{m} f_j}. \qquad (4.15)$$

Користуючись правилом розкладання дисперсій, можна за двома відомими дисперсіями знайти третю – невідому, а також мати уяву про силу впливу групувальної ознаки.

Контрольні запитання

1. Поняття та види статистичних показників.
2. Поняття та види абсолютних величин.
3. Поняття та види відносних величин.
4. Поняття та види середніх величин.
5. Середня гармонійна та формули для її обчислення.
6. Середня квадратична та формули для її обчислення.
7. Середня геометрична та формули для її обчислення.
8. Мода та медіана ряду розподілу.
9. Квартилі та децилі для рядів розподілу.
10. Суть варіації.
11. Види основних показників варіації.
12. Правило розкладання дисперсії та складові загальної дисперсії.

5. ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ

5.1. Поняття про вибіркове спостереження

При статистичному обстеженні різних явищ суспільного життя часто доводиться зустрічатися з прикладами недоцільності або неможливості проведення *суцільного спостереження,* тобто вивчення всіх одиниць сукупності. Так, *недоцільно* проводити обстеження бюджетів сімей в обсязі всієї країни, оскільки це було б пов'язане із залученням тисяч статистиків та значними матеріальними витратами. Практично *неможливо* на підприємстві для контролю якості хлібобулочних виробів, консервів тощо проводити суцільний контроль, оскільки це призведе до пошкодження або знищення всієї партії продуктів. Тому у випадках, коли суцільне спостереження є *недоцільним,* використовують *несуцільне спостереження,* різновидом якого є *вибірка*. Цей вид спостереження широко використовується в соціологічних дослідженнях, обстеженні якості продуктів харчування, обстеженні домогосподарств, маркетингових дослідженнях, аудиторських перевірках тощо. Крім того, вибірковий метод використовується для прискорення обробки матеріалів суцільного спостереження, перевірки правильності даних переписів, проведення спостережень.

Використання вибіркового методу замість суцільного спостереження дає можливість зберігати трудові та матеріальні ресурси і кошти, провести спостереження в стислі строки та отримати кінцеві результати в більш короткі терміни часу.

Вибірковий метод використовується для опису явищ при використанні закону великих чисел.

Всі одиниці явища називаються *генеральною сукупністю,* а окрема частина цих одиниць, відібраних із генеральної сукупності для безпосереднього спостереження, називається *вибірковою сукупністю.* Кажуть, що вибіркова сукупність *репрезентує* (представляє) всю генеральну сукупність.

Об'єктивну гарантію репрезентативності отриманої вибірки дає використання відповідних науково обґрунтованих *способів відбору* одиниць вибіркової сукупності:

- вибірка з генеральної сукупності має бути проведена *випадково,* тобто кожна її одиниця повинна мати таку ж ймовірність потрапити у вибірку (так, наприклад, відібрані найкращі або найгірші одиниці не відображають дійсний розподіл ознаки в генеральній сукупності);
- вибірка має бути здійснена із *однорідної* сукупності, оскільки за інших обставин результати вибірки будуть не точними і не можуть в повній мірі репрезентувати генеральну сукупність.

При створенні випадкової вибірки можливі *два підходи:*

1) *відбір при жеребкуванні* заздалегідь занумерованих одиниць генеральної сукупності;
2) використання *таблиць випадкових чисел.*

У першому підході розрізняють два принципово різних способи формування вибіркової сукупності:

- *повторна вибірка,* коли відібрана з генеральної сукупності занумерована одиниця фіксується і знов повертається на своє місце, після чого пачка номерів одиниць генеральної сукупності ретельно перемішується; цей спосіб відбору на практиці є обмеженим через недоцільність, а іноді й неможливість повторного обстеження;
- *безповторна вибірка,* коли відібраний із пачки номер одиниці генеральної сукупності відкладається в сторону і не повертається назад у пачку; цей спосіб відбору характеризується підвищеним ступенем точності, надійності вибірки і найчастіше використовується на практиці.

При *другому підході* із таблиці випадкових чисел відбирають n чисел із будь – якого рядка або стовпця таблиці, кількість яких не перевищує N чисел генеральної сукупності; потім відбирають будь-яким способом ті одиниці заздалегідь занумерованої сукупності із n чисел, які відповідають відібраним числам таблиці, що і складає вибіркову сукупність.

У статистичній практиці розрізняють такі *різновиди* вибірки:

- за способом організації вибіркового обстеження;
- за ступенем охоплення одиниць обстежуваної сукупності.

За *способом організації* використовують такі *види* вибірки:

- проста випадкова вибірка;

- механічна вибірка;
- районована (типова) вибірка;
- серійна вибірка;
- ступенева вибірка.

За *ступенем охоплення одиниць обстежуваної* сукупності вибірки бувають:

1) великі (при $n \geq 30$);
2) малі (при $n < 30$).

5.2. Характеристики генеральної та вибіркової сукупності

Нехай нас цікавить ознака x обсягом N одиниць в генеральній сукупності, що представляється таким варіаційним рядом 1 (табл. 5.1).

Таблиця 5.1 – Розподіл одиниць генеральної сукупності

Варіанти x	x_1	x_2	…	x_i	…	x_M	x
Частоти F	F_1	F_2	…	F_i	…	F_M	$\sum_{i=1}^{M} F_i = N$

Цей розподіл невідомий, бо якщо б ми його знали, то відпала б необхідність в організації вибірки.

Узагальнюючими характеристиками цього ряду 1 будуть такі:

1) *генеральна середня* $\bar{x}$

$$\bar{x} = \frac{\sum x_i F_i}{N}; \tag{5.1}$$

2) *генеральна дисперсія* σ^2:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 F_i}{N}; \tag{5.2}$$

3) *генеральне середнє квадратичне відхилення* σ;

4) *частка ознаки одиниць генеральної сукупності p*, тобто частка одиниць M, яка володіє даним значенням ознаки у загальному обсязі N генеральної сукупності

$$p = M/N. \tag{5.3}$$

Мета вибіркового обстеження полягає в тому, щоб, відібравши з генеральної сукупності п одиниць, обстежити їх і на цій основі оцінити невідомі нам генеральні характеристики. Варіація ознаки *x* у вибірковій сукупності обсягом *n* може бути представлена у вигляді варіаційного ряду 2 (табл. 5.2):

Таблиця 5.2 – Розподіл одиниць вибіркової сукупності

Варіанти *x*	x_1	x_2	…	x_i	…	x_m	x
Частоти *f*	f_1	f_2	…	f_i	…	f_m	$\sum_{i=1}^{m} f_i = n$

Узагальнюючими характеристиками ряду 2 вибіркової сукупності будуть:

1) *вибіркова середня* $\tilde{x}$

$$\tilde{x} = \frac{\sum x_i f_i}{n}; \qquad (5.4)$$

2) *вибіркова дисперсія* $\sigma_{\text{в}}^2$

$$\sigma_{\text{в}}^2 = \frac{\sum (x_i - \tilde{x})^2 f_i}{n}; \qquad (5.5)$$

3) *вибіркове середнє квадратичне відхилення* $\sigma_{\text{в}}$;

4) *частка ознаки одиниць вибіркової сукупності w,* тобто відношення кількості одиниць вибіркової сукупності *m,* яка володіє даною ознакою, до обсягу вибіркової сукупності *n*

$$w = m/n; \qquad (5.6)$$

5) *частка вибірки* $w_{\text{в}}$, як відношення обсягу вибірки до обсягу генеральної сукупності $w_{\text{в}} = n/N$.

5.3. Помилки вибіркового спостереження

При правильному проведенні вибіркового спостереження характеристики вибірки близькі до відповідних характеристик генеральної сукупності, але все ж таки вони не збігаються. Пояснюється це наявністю помилки вибірки. *Помилкою вибірки* називаються деякі розходження характеристик генеральної та вибіркової сукупностей. Вона складається із помилок реєстрації та помилок репрезентативності.

Помилками реєстрації називають такі, які виникають внаслідок отримання неточних або неправільних відомостей від окремих одиниць сукупності із-за недосконалості вимірювальних приладів, недостатньої кваліфікації спостерігача, недостатньої точності розрахунку тощо. Ці помилки повинні бути виключені або зведені до мінімуму.

Помилки репрезентативності поділяють на систематичні та випадкові. *Систематичні помилки репрезентативності* виникають внаслідок особливостей прийнятої системи та обробки даних спостереження або черерз недотримання правил відбору у вибіркову сукупність. Такі помилки також повинні бути виключені. *Випадкові помилки репрезентативності* виникають перш за все через те, що вибіркова сукупність через її малий обсяг не завжди точно відтворює характеристики генеральної сукупності. Тому цей вид помилок вибірки є основним, і завдання вибіркового методу полягає в отриманні таких вибіркових характеристик, які б якомога точніше відтворювали характеристики генеральної сукупності, тобто давали найменші помилки репрезентативності.

Теорія вибіркового методу полягає в знаходженні середньої величини помилки репрезентативності та можливих їх меж при різних способах утворювання вибіркової сукупності. Для кожного конкретного вибіркового спостереження значення помилки репрезентативності обчислюється за відповідними формулами, які будуть розглянуті нижче.

5.4. Закон великих чисел

При вибірковому обстеженні повинна бути забезпечена випадковість відбору одиниць у вибіркову сукупність з метою достатньої репрезентативності генеральної сукупності. Тому вибірковий метод спостереження основано на ймовірному підході, теоретичною базою для якого є *закон великих чисел.* Він детально розглядався в курсі теорії ймовірностей і математичної статистики. Але, враховуючи його принципове значення при вибіркових обстеженнях, нагадаємо головні положення цього закону, який складає математичну основу вибіркового методу.

Суть закону великих чисел полягає в тому, що при збільшенні чисельності одиниць сукупності поступово зменшується елемент випадковості в узагальнених характеристиках сукупності. На основі закону можна стверджувати, що при достатньо великому обсязі вибірки ($n \geq 30$) вибіркові характеристики мало відрізняються від генеральних, тому використовуються наближені рівняння для середньої, частки, дисперсії, середнього квадратичного відхилення

$$\bar{x} \approx \tilde{x}\,;\; p \approx \omega;\; \sigma \approx \sigma_{\text{в}}. \tag{5.7}$$

В теорії ймовірностей закон великих чисел виражає ряд теорем.

Так, *теорема Чебишева,* з якої як окремі випадки випливають теореми Бернуллі та Пуассона, стверджує, що при необмеженому збільшенні кількості незалежних спостережень $n \to \infty$ в генеральній сукупності при обмеженій дисперсії з ймовірністю, скільки завгодно наближеною до одиниці, можна стверджувати, що вибіркові характеристики (середня, частка) будуть скільки завгодно мало відрізнятися від відповідних генеральних характеристик, тобто

$$P\left(\left|\tilde{x} - \bar{x}\right| < \varepsilon\right) \to 1 \text{ при } n \to \infty, \tag{5.8}$$

де P – ймовірність нерівності у круглих дужках; ε – будь-яке завгодно мале додатне число; $\tilde{x}, \bar{x}$ – вибіркова та генеральна середні.

5.5. Проста випадкова вибірка

Поняття і категорії, які лежать в основі простої випадкової вибірки, є вихідними при розробці інших видів вибіркового спостереження. Проста випадкова вибірка є однією з найпоширених видів відбору із генеральної сукупності.

При *простій випадковій вибірці* відбір одиниць здійснюється із всієї маси одиниць генеральної сукупності без попереднього розподілення її на будь-які групи і одиниці відбору співпадають з одиницями обстеження.

Як зазначалось, з практичної точки зору перевага віддається простій безповторній вибірці, яка може формуватися на основі жеребкування одиниць сукупності або при використанні таблиць випадкових чисел (їх можуть замінити таблиці логарифмів).

Необхідно особливо підкреслити, що важливою умовою репрезентативності випадкового відбору є те, що кожній одиниці генеральної сукупності надається однакова можливість потрапити у вибіркову сукупність. Саме принцип випадковості попадання будь-якої одиниці генеральної сукупності у вибірку запобігає виникненню систематичних помилок відбору.

Одним із прикладів використання простої випадкової вибірки є проведення тиражів виграшів грошово-речової лотереї, при якій забезпечується однакова можливість попадання в тираж будь-якого номера лотерейного квитка.

При простій випадковій вибірці (як і в інших видах вибіркового спостереження) можливо рішення таких *задач:*

- визначення *помилки* вибіркового спостереження;
- визначення *меж генеральних характеристик* на основі вибіркових із заданою довірчою ймовірністю (ступенем надійності);
- визначення *довірчої ймовірності* тому, що генеральні характеристики можуть відрізнятися від вибіркових не більше певної заданої величини;
- знаходження необхідної *чисельності вибірки,* яка б з практичною достовірністю забезпечувала задану точність вибіркових характеристик.

Розв'язання цих задач може проводитися як по відношенню до генеральної середньої арифметичної $\bar{x}$, так і до частки $\bar{p}$. Розглянемо перераховані задачі у відповідності *до безповторної вибірки*, яка на практиці зустрічається найбільш часто.

При рішенні *першої задачі* в математичній статистиці доводиться, що при великій кількості одиниць вибіркової сукупності ($n \geq 30$) *середня квадратична помилка* безповторної вибірки μ визначається за формулами

а) для середньої $$\mu_x = \sqrt{\frac{\sigma_{\text{в}}^2}{n}\left(1 - \frac{n}{N}\right)}; \qquad (5.7)$$

б) для частки $$\mu_w = \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)}. \qquad (5.8)$$

На основі теореми Ляпунова *гранична помилка вибірки* $\Delta = t\mu$. *Коефіцієнт довіри* t при визначенні граничної помилки залежить від прийнятого рівня ймовірності P: так, при t = 1,0 значення ймовірності P = 0,683; t = 1,96 – для ймовірності P = 0,950; t = 2,0 – для ймовірності P = 0,954; t = 3,0 – для ймовірності P = 0,997.

Одним з основних напрямків дослідження при використанні вибіркового методу є *оцінка за даними вибірки характеристик генеральної сукупності*, що відноситься до можливостей зазначеної *другої задачі.*

Величини генеральної середньої та частки можуть бути представлені *інтервальною оцінкою* у вигляді визначення *довірчого інтервалу* із заданого рівня *довірчої ймовірності P*:

а) *для середньої* $$\tilde{x} - \Delta_x \leq \bar{x} \leq \tilde{x} + \Delta_x; \qquad (5.9)$$

б) *для частки* $$w - \Delta_w \leq p \leq w + \Delta_w. \qquad (5.10)$$

Формули (5.9) і (5.10) встановлюють межі, в яких при заданій довірчій ймовірності знаходиться невідома величина оцінюваного параметра: середньої x або частки p в генеральній сукупності. Ймовірність того, що величина генеральної середньої або частки вийде за довірчі межі, дорівнює $\alpha = 1 - P$ і називається *рівнем значущості (істотності).* Для ймовірності P = 0,950 або P = = 0,954 рівень значущості дорівнює відповідно 0,050 (або 5,0 %) та 0,046 (або

4,6 %), і перевищення меж у довірчих інтервалах (5.9), (5.10), що має таку ймовірність, практично неможливе.

Іноді доводиться вирішувати *третю задачу,* коли необхідно визначити *довірчу ймовірність* того, що генеральні характеристики відрізняються від вибіркових не більше заданої величини *P*.

Довірча ймовірність *P*, яку необхідно обчислити за теоремою Ляпунова, є функцією від коефіцієнта t: $P = \Phi(t)$, де $\Phi(t)$ – інтеграл Лапласа $\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$. Значення t у свою чергу може бути визначено через граничну та стандартну помилки $t = \Delta / \mu$, обчислені відносно середньої або частки. Нарешті, за знайденим значенням t із довідкових таблиць розраховується інтеграл Лапласа, який відповідає обчислюваній ймовірності *P*, яка порівнюється із заданою величиною.

Однією із основних задач вибіркового методу є визначення *чисельності вибірки n (четверта задача* в нашій класифікації). У випадку безповторного відбору чисельність вибірки обчислюється за формулами

а) для середньої $$n = \frac{t^2 \sigma_{\text{в}}^2 N}{\Delta_x^2 N + t^2 \sigma_{\text{в}}^2}; \quad (5.11)$$

б) для частки $$n = \frac{t^2 w(1-w) N}{\Delta_w^2 N + t^2 w(1-w)}. \quad (5.12)$$

Область застосування простої випадкової вибірки надзвичайно широка: перевірки різних одиниць сукупностей; численні обстеження підприємств, установ, їхніх працюючих, населення; дослідження в сільськогосподарських задачах (якості продукції, польових дослідженнях, визначення втрат урожаю тощо).

5.6. Механічна вибірка

Механічною називається така вибірка, при якій генеральна сукупність обсягом *N* одиниць, розташованих у певному порядку (за зростанням або зменшенням, за алфавітом, географічним положенням тощо), розділяється на *n* рівних частин і з кожної частини обстежується одна одиниця.

Відношення *N/n* – називається *інтервалом* вибірки. Наприклад, якщо відбір складає 5 % від генеральної сукупності працюючих на підприємстві, розміщених у списку за алфавітним порядком, то обстежують кожного 20-го працюючого (5 % – це 1/20 спискового складу працюючих). Інтервал вибірки буде дорівнювати 100/5 = 20 %. За початок відрахунку при обстеженні генеральної сукупності в списках приймають або *початкову одиницю,* визначену випадковим відбором (при невпорядкованому розміщенні одиниць генеральної сукупності), або *середину першого інтервалу* (якщо одиниці в списку розміщені за певною ознакою – зростанням або збільшенням). Механічна вибірка дуже зручна у випадках, коли вже списки одиниць, складені в тому чи іншому порядку, або тоді, коли ми не можемо заздалегідь скласти список одиниць генеральної сукупності і які з'являються поступово протягом якогось періоду (наприклад: під час оплати покупок в магазині обстежити кожного 10-го покупця; при контролі якості продукції – перевірити кожну 5-ту деталь, яка зійшла зі станка).

Помилки вибірки при механічному відборі одиниць обчислюють за формулами простої випадкової безповторної вибірки.

З метою економії часу та засобів іноді буває зручно обстежувати не всю вибіркову сукупність, а частину її, тобто здійснити *підвибірку* з одиниць первісної вибірки. Цей спосіб називають *двофазним,* а при наявності декількох підвибірок – *багатофазовим.* Останній спосіб найчастіше використовують у тих випадках, коли кількість необхідних для визначення показників має різну точність (наприклад, у випадках різного ступеня варіації показників). Помилки при багатофазовій вибірці розраховуються у кожній фазі окремо.

Іноді буває доцільним взяти з сукупності дві або більше незалежних між собою вибірок, використовуючи для кожної з них однаковий спосіб відбору. Такі вибірки називають *взаємнопроникнутими вибірками.* Перевага таких вибірок полягає в тому, що вони дозволяють отримати окремі і незалежні оцінки тих або інших ознак сукупності.

5.7. Районована (типова) вибірка

Районованою вибіркою називають такий спосіб відбору, який здійснюється на основі розподілу кількості відібраних одиниць n між районами (групами), які є в генеральній сукупності. Районами залежно від характеру генеральної сукупності, можуть бути територіальні області, галузі виробництва, окремі підприємства, соціальні групи населення тощо. Якщо генеральна сукупність розділяється на m частин, груп, районів тощо, тобто $N = N_1+N_2+\ldots+N_i+N_m$, то і вибіркова сукупність повинна формуватися із m частин так, щоб $n = n_1+n_2+\ldots+n_i+\ldots+n_m$. При цьому розподіл між районами може бути різним:

а) *пропорційним,* коли кількість одиниць, які відбираємо у вибірку, є пропорційною до питомої ваги району в генеральній сукупності, тобто кількість спостережень у кожному районі розраховується за формулою

$$n_i = n\frac{N_i}{N}; \qquad (5.13)$$

б) *непропорційним*, якщо з кожного району відбирають однакову кількість одиниць

$$n_i = \frac{n}{k}; \qquad (5.14)$$

де k – кількість відокремлюваних районів;

в) *оптимальним*, яке враховує і чисельність району N_i і середнє квадратичне відхилення ознаки в районі σ_i; тоді чисельність кожного району вибірки n_i розраховується за формулою

$$n_i = n\frac{\sigma_i N_i}{\sum \sigma_i N_i}. \qquad (5.15)$$

На практиці в більшості випадків застосовують перший і третій способи розподілення між районами. Але використання оптимального розміщення ускладнюється тим, що ми не завжди маємо дані про величини σ_i в генеральній сукупності. Тому в таких випадках вибираємо найбільш часто використовуваний пропорційний розподіл між районами. Наведемо для цього способу розподілення формули розрахунку *середньої квадратичної помилки вибірки* при безповторному відборі усередині районів

а) *для середньої*

$$\mu_x = \sqrt{\frac{\overline{\sigma}_{\text{в}}^2}{n}\left(1 - \frac{n}{N}\right)}, \tag{5.16}$$

де $\overline{\sigma}_{\text{в}}^2$ – середня з дисперсій районів вибірки $\overline{\sigma}_{\text{в}}^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i}$;

б) *для частки*

$$\mu_w = \sqrt{\frac{\overline{w(1-w)}}{n}\left(1 - \frac{n}{N}\right)}, \tag{5.17}$$

де $\overline{w(1-w)}$ – середня з часток районів.

Визначення необхідної чисельності вибірки при безповторному відборі усередині районів здійснюється за формулами

а) для середньої

$$n = \frac{t^2 \overline{\sigma}_{\text{в}}^2 N}{\Delta^2 N + t^2 \overline{\sigma}_{\text{в}}^2}; \tag{5.18}$$

б) для частки

$$n = \frac{t^2\, \overline{w(1-w)}}{\Delta^2 N + t^2\, \overline{w(1-w)}}. \tag{5.19}$$

5.8. Серійна вибірка

Суть *серійної вибірки* полягає в тому, що відбору підлягають окремі серії (групи, гнізда) одиниць генеральної сукупності. На практиці часто зустрічається відбір з рівними серіями. У відібраних методом випадкового безповторного або механічного відбору серіях проводять суцільне спостереження всіх одиниць, що до них увійшли.

Серійна вибірка має переваги організаційного характеру у порівнянні з іншими видами відбору і широко використовується там, де генеральна сукупність складається з певним чином відокремлених груп (наприклад, у випадку статистичного контролю якості готової продукції, упакованої в пачки, ящики, контейнери для транспортування, зберігання та продажу: зручніше перевірити декілька упаковок-серій, ніж із усіх упаковок відібрати необхідну кількість товару).

Оскільки при серійній вибірці кожна серія виступає як самостійна одиниця спостереження, то дисперсія усередині серій у випадку визначення середньої помилки та чисельності вибірки має бути виключена і враховується тільки міжсерійна дисперсія δ^2.

При рівних серіях *середня квадратична помилка* безповторної вибірки та її *чисельність* обчислюються за формулами

$$\mu = \sqrt{\frac{\delta^2}{r}\left(1 - \frac{r}{R}\right)}; \tag{5.20}$$

$$r = \frac{t^2 \delta^2 R}{R\Delta^2 + t^2 \delta^2}, \tag{5.21}$$

де *r* – кількість відібраних серій; *R* – загальна кількість серій в генеральній сукупності. При цьому міжсерійна дисперсія розраховується так:

а) для середньої $$\delta^2 = \frac{\sum\left(\tilde{x}_i - \tilde{x}_0\right)^2}{r}; \tag{5.22}$$

б) для частки $$\delta^2 = \frac{\sum\left(w_i - \overline{w}\right)^2}{r}, \tag{5.23}$$

де $\tilde{x}_i$ – середні в серіях, $\tilde{x}_0$ – загальна середня для серій $\tilde{x}_0 = \frac{\sum x_i}{r}$; w_i – частки в серіях (групах); $\overline{w}$ – середня частка ознаки для всієї вибіркової сукупності.

Чим менше групові середні і частки відрізняються одна від одної, тобто чим ближче одна від одної серії за рівнем прийнятої ознаки, тим точніше серійна вибірка.

5.9. Ступенева вибірка

Серійну вибірку можна розглядати як так звану *одноступеневу вибірку,* де у випадково відібраних серіях генеральної сукупності проводять суцільний опис усіх одиниць, що до них включені. Але можливо сформувати вибіркову сукупність у два етапи: на першому етапі методом випадкового безповторного відбору формують серії, які підлягають обстеженню; на другому етапі в кожній серії випадковим безповторним відбором формується певна кількість одиниць для подальшого обстеження. *Середня квадратична помилка* вибірки буде залежати від помилки серійного відбору та помилки індивідуального відбору.

5.10. Малі вибірки

Розглянуті вище види вибірок вважали великими за кількістю одиниць обстеження ($n \geq 30$). Але на практиці часто зустрічаються і з *малими вибірками* ($n<30$): наприклад, у технічному нормуванні; агрономічних та зоотехнічних дослідженнях; контролі якості продукції, пов'язаному із знищенням зразків; вибіркових фотографіях робочого дня тощо. В таких випадках для розрахунку помилки вибірки неможливо користуватися теоремами закону великих чисел, оскільки на вибіркову середню великий вплив має значення кожної із випадково відібраних одиниць і їх розподіл може значно відрізнятися від нормального закону розподілу. Тому при вибірках невеликого обсягу методи оцінювання результатів вибіркового спостереження мають свою особливість порівняно з методами великих вибірок.

Теоретичні положення для оцінки характеристик малих вибірок вперше розробив англійський статистик В. Госсет (1908 р.), який друкував свої праці під псевдонімом Ст'юдента. Пізніше теоретичні питання малих вибірок були розвинуті Р. Фішером (1925 р.) та іншими.

Припускаючи, що вибірки зроблено з нормально розподіленої генеральної сукупності, Ст'юдент встановив закон розподілу відхилень вибіркових характеристик від генеральних для малих вибірок (відкритий ним розподіл дістав назву *t-розподілу Ст'юдента*, що подібний до нормального закону).

Відхилення вибіркової середньої $\tilde{x}$ від генеральної середньої $\bar{x}$ Ст'юдент виразив у вигляді *відношення Ст'юдента*. Фактично це коефіцієнт довіри між граничною $\Delta_{\text{м.в}}$ та середньою квадратичною $\mu_{\text{м.в}}$ помилками у малій вибірці $\Delta_{\text{м.в}} = \mu\ t_{\text{м.в}}$

$$t = \frac{\tilde{x} - \bar{x}}{\mu_{\text{м.в}}}. \tag{5.24}$$

Значення t може бути знайдено за математичними таблицями розподілу Ст'юдента залежно від рівня значущості $\alpha = 1 - P$ (P – рівень ймовірності числа степенів вільності варіації $k = n–1$ (n – обсяг малої вибірки).

Середня квадратична помилка для кількостей ознак: малої вибірки $\mu_{\text{м.в}}$ визначається за формулою

$$\mu_{\text{м.в}} = \sqrt{\frac{\sigma^2_{\text{м.в}}}{n}}, \tag{5.25}$$

де $\sigma^2_{\text{м.в}}$ – дисперсія малої вибірки

$$\sigma^2_{\text{м.в}} = \frac{\sum_{i=1}^{n}\left(x_i - \tilde{x}\right)^2}{n-1}. \tag{5.26}$$

Ймовірність того, що помилка вибірки буде не більше заданого значення $|\tilde{x} - \bar{x}| \le t\mu_{\text{м.в}}$, *представляє собою функцію* $S(t,n)$, *наведену в таблицях Ст'юдента.*

$$S(t,n) = P(|\tilde{x} - \bar{x}| \le t\mu_{\text{м.в}}). \tag{5.27}$$

Із таблиць Ст'юдента слідує, що при збільшенні обсягу вибірки розподіл Ст'юдента наближується до нормального закону і при $n = 20$ він мало відрізняється від нормального розподілу.

Слід врахувати, що розподіл Ст'юдента слід використовувати тільки для оцінки помилок вибірки, взятої із генеральної сукупності з нормальним законом розподілення ознаки.

5.11. Поняття про метод моментних спостережень

При вибірковому спостереженні ми розглядали сукупність одиниць, яка існує незмінно або поступово формується протягом деякого періоду. Але вибіркове спостереження може зводитися до фіксації стану безперервного процесу на певні моменти часу і, отримавши назву *моментних спостережень*, використовується у статистичній практиці при вивченні технологічних процесів (структури витрат робочого часу різних категорій працюючих, характеристики використання виробничого устаткування тощо).

Суть методу полягає у фіксації стану процесу або виду витрат часу у певні моменти. Складається перелік усіх можливих станів процесу або витрат часу. Після закінчення спостереження дослідник встановлює частку відміток за кожним станом або видом витрат часу в загальному обсязі спостережень. Частка часу, витраченого на даний вид роботи, може бути оцінена за допомогою частки моментів, коли виконувалася ця робота, в загальній кількості спостережень.

Середня помилка частки w визначається як для простої повторної випадкової вибірки за формулою

$$\mu_w = \sqrt{\frac{w(1-w)}{n}}. \tag{5.28}$$

Гранична помилка вибірки $\Delta = \mu\ t_{\text{м.в}}$ розраховується з використанням значення коефіцієнта довіри t, який визначається прийнятим рівнем довірчої ймовірності P (наприклад, при $P = 0{,}954$ $t = 2{,}0$).

Відбір моментів може бути організовано за схемою механічної вибірки через рівні інтервали часу або за схемою випадкової вибірки з використанням таблиць випадкових чисел.

Для визначення *чисельності* моментних спостережень рекомендується формула повторної випадкової вибірки, де не враховується чисельність одиниць N генеральної сукупності

$$n = \frac{\sigma^2 t^2}{\Delta^2}.$$

Хоча вибірка у моментних спостереженнях є безповторною, визначити чисельність одиниць генеральної сукупності неможливо (вона фактично безмежна при моментах спостережень малої тривалості). Крім того, дисперсія σ^2 ознаки у генеральній сукупності здебільшого також невідома і може бути знайдена тільки після проведення вибіркового спостереження. Тому приймають максимальне значення дисперсії $\sigma^2 = 0{,}25$. Якщо довірча ймовірність $P = 0{,}954$, то при $t = 2{,}0$ чисельність вибірки за методом моментних спостережень визначиться за формулою: $n = 0{,}25 \cdot \dfrac{2^2}{\Delta^2} = \dfrac{1}{\Delta^2}$.

Контрольні запитання

1. Поняття вибіркового спостереження.
2. Види вибірки в статистичних дослідженнях.
3. Узагальнюючі характеристики в генеральній і вибірковій сукупностях.
4. Помилки вибіркового спостереження та їх види.
5. Суть закону великих чисел.
6. Суть простої випадкової вибірки.
7. Суть механічної вибірки.
8. Суть районованої (типової) вибірки.
9. Суть серійної та ступеневої вибірки.
10. Особливості малої вибірки у порівнянні з великою.
11. Поняття про метод моментних спостережень.

6. СТАТИСТИЧНІ МЕТОДИ АНАЛІЗУ КОРЕЛЯЦІЙНИХ ЗВ'ЯЗКІВ

6.1. Види зв'язку між ознаками явищ

Усі соціально-економічні явища взаємопов'язані та взаємозумовлені і зв'язок (залежність) між ними носить причинно-наслідковий характер. Суть причинного зв'язку полягає в тому, що при необхідних умовах одне явище зумовлює інше і в результаті такої взаємодії виникає наслідок.

Особливу актуальність має вивчення взаємозв'язку в умовах ринкової економіки і є важливою функцією діяльності менеджерів та економістів. Вивчення механізму ринкових зв'язків, взаємодії попиту і пропозиції, вплив обсягу і складу пропозиції товарів на обсяг і структуру товарообороту, формування товарних запасів, прибутку та інших якісних показників має першочергове значення для прогнозування кон'юнктури ринку та вирішення багатьох питань успішного ведення бізнесу.

Вивчаючи закономірності зв'язку, причини і умови, що їх характеризують, об'єднують у поняття *фактора*. Тоді ознаки, що є причинами та умовами зв'язку, називаються *факторними* x, а ті, що змінюються під впливом факторних ознак, – *результативними* y.

Між ознаками x та y існують різні за природою та характером *види зв'язку:* функціональні та стохастичні.

При *функціональному зв'язку* між факторною та результативною ознаками кожному значенню ознаки x відповідає одне чітко визначене значення ознаки y. Такі зв'язки найчастіше вивчаються в математичному аналізі і використовуються для встановлення кількісних співвідношень у точних та прикладних науках (математиці, фізиці, астрономії тощо). Прикладом функціонального зв'язку може бути залежність між радіусом кола R (факторна ознака x) та довжиною кола C (результативна ознака y) у формулі $C = 2\pi R$, де кожному значенню радіуса R відповідає одне конкретне значення довжини кола C. Зазначимо, що функціональні зв'язки між ознаками вивчаються в економіці за допомогою індексного методу.

При *стохастичному зв'язку* кожному окремому значенню факторної ознаки x відповідає певна множина значень результативної ознаки y. Такий зв'язок утворює *умовний розподіл* ознак, який варіює. Наприклад відомо, що урожайність залежить від кількості внесених добрив. Але на урожайність впливає ще багато інших факторів (строки внесення добрив, глибина їх внесення тощо). Зв'язки такого виду називають ще *статистичними, ймовірними*.

Підвидом стохастичного зв'язку є *кореляційна залежність,* що зумовлює кореляційний зв'язок між ознаками. При такій залежності зі зміною факторної ознаки x змінюються групові середні результативної ознаки y і замість умовних розподілів множин значень ознаки y виступають середні значення цих розподілів. Таким чином, між ознаками x та y існує кореляційна залежність, коли середня величина однієї з них змінюється залежно від значення іншої.

Прикладом вихідних даних для встановлення кореляційного зв'язку між змінними x та y може бути дискретний розподіл, який характеризується *кореляційною таблицею*.

Визначення кореляційного зв'язку між ознаками займає значне місце у дослідженнях соціально-економічних явищ в економіці і управлінні. Зміст такого зв'язку складає *теорія кореляції.* Основоположниками цієї теорії є англійські вчені-біологи Ф. Гамільтон (1822 – 1911 рр.), К. Пірсон (1857 – 1936 рр.). Термін *"кореляція"* взято із природознавства і означає співвідношення, відповідність між змінними у рівнянні регресії.

Умовами правильного використання методів теорії кореляції є такі:

а) наявність *однорідності* тих одиниць, які підлягають дослідженню (наприклад, відбір підприємств, які випускають однотипну продукцію, мають однаковий характер технології і тип обладнання тощо);

б) достатньо *велика кількість спостережень,* при яких ми погашаємо вплив випадковостей на результативну ознаку і має силу закон великих чисел;

в) *нормальний характер розподілу* результативної ознаки, на якому побудовані всі положення теорії кореляції.

В основі теорії кореляції лежить *кореляційно-регресійний аналіз* (*КРА*), суть якого полягає у виборі виду рівняння регресії, обчисленні його параметрів та встановленні адекватності (відповідності) теоретичної залежності фактичним даним. Наявність такої теоретичної залежності значно полегшує аналіз економічних явищ, дає змогу встановлення прогнозу на майбутнє. Докладно питання застосування КРА, меж його використання та характеристика інших методів обстеження зв'язку між ознаками економічних явищ будуть розглянуті під час подальшого вивчення дисципліни "Економетрія". Тому в даному розділі зупинимося на принциповій сутності питань КРА при використанні його у простіших випадках.

6.2. Види рівнянь регресії та визначення їх параметрів

Якщо змінна y залежить від однієї змінної, то рівняння регресії є найпростішим і має назву *рівняння парної регресії*. Якщо y залежить від більше ніж однієї незалежної змінної, то така залежність має назву рівняння *множинної або багатофакторної регресії* (табл. 6.1).

Таблиця 6.1 – Види залежностей та рівнянь регресії

Вид залежності	Парна регресія	Множинна регресія
лінійна	$\hat{y} = a_0 + a_1 x$	$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_m x_m$
квадратична	$\hat{y} = a_0 + a_1 x^2$	…
гіперболічна	$\hat{y} = a_0 + a_1 / x$	…
степенева	$\hat{y} = a_0 x^{a_1}$	…
логарифмічна	$\hat{y} = a_0 + a_1 \ln x$	…

Метод найменших квадратів (МНК) полягає у такому: невідомі параметри a_j в табл. 6.1 обираються таким чином, щоб сума квадратів відхилень емпіричних (фактичних) значень y_i від розрахункових $\hat{y}_i$ була мінімальною

$$S = \sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2 \to \min . \qquad (6.1)$$

Необхідна умова екстремуму функції

$$\frac{\partial S}{\partial a_j} = 0,\ j=1..p, \qquad (6.2)$$

де p – кількість параметрів у системі.

Формули (6.1), (6.2) – це загальні формули МНК, які застосовуються для знаходження параметрів будь-якого регресійного рівняння з будь-якою кількістю незалежних змінних.

6.3. Лінійна парна регресія

Рівняння лінійної парної регресії має вигляд

$$\hat{y} = b_0 + b_1 x. \qquad (6.3)$$

Сума квадратів для парної лінійної регресії відповідно до (6.1), (6.2) матиме такий вигляд: $S = \sum_{i=1}^{n} (b_0 + b_1 x_i - y_i)^2$.

Прирівняємо до нуля її частинні похідні

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = 2\sum_{i=1}^{n} (b_0 + b_1 x_i - y_i) = 0; \\ \dfrac{\partial S}{\partial b_1} = 2\sum_{i=1}^{n} (b_0 + b_1 x_i - y_i) x_i = 0, \end{cases} \qquad (6.4)$$

звідки після перетворень отримаємо систему нормальних рівнянь для визначення параметрів лінійної системи

$$\begin{cases} b_0 n + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i; \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases} \qquad (6.5)$$

Тепер, розділивши обидві частини рівняння (6.5) на n, отримаємо систему нормальних рівнянь у вигляді

$$\begin{cases} b_0 + b_1 \bar{x} = \bar{y}; \\ b_0 \bar{x} + b_1 \bar{x}^2 = \overline{xy}, \end{cases} \qquad (6.6)$$

де відповідні середні визначаються за формулами

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}; \qquad \overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n};$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}; \qquad \overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n}.$$

Підставляючи значення $b_0 = \bar{y} - b_1\bar{x}$ з першого рівняння системи (6.6) у рівняння регресії (6.3), отримаємо

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{K_{XY}}{s_x^2}, \qquad (6.7)$$

де b_1 – вибірковий коефіцієнт регресії, K_{XY} – вибірковий кореляційний момент або вибіркова кореляція, s_x^2 – вибіркова дисперсія змінної X

$$K_{XY} = \overline{xy} - \bar{x} \cdot \bar{y}, \; s_x^2 = \overline{x^2} - \bar{x}^2.$$

Величина $r = \frac{b_1 \cdot s_x}{s_y}$ називається *вибірковим коефіцієнтом кореляції*. Він показує, наскільки величина s_y зміниться Y, якщо X зміниться на одне s_x.

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x \cdot s_y}. \qquad (6.8)$$

Підставляючи у вираз (6.8) вихідні дані, отримаємо

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}. \qquad (6.9)$$

Коефіцієнт кореляції має такі властивості:

1) Коефіцієнт кореляції приймає значення на відрізку [-1;1], тобто $-1 \le r \le 1$. Чим ближче $|r|$ до 1, тим тіснішим є кореляційний зв’язок.

2) При $|r| = 1$ кореляційний зв’язок стає функціональним. При цьому всі значення, що спостерігаються, лежать на одній лінії.

3) При $r = 0$ кореляційний зв’язок відсутній і лінія регресії паралельна осі x.

При $r > 0$ ($b_1 > 0$) кореляційний зв’язок називають *прямим*.

При $r < 0$ ($b_1 < 0$) кореляційний зв’язок називають *оберненим*.

Контрольні запитання

1. Поняття факторних та результативних ознак.
2. Види зв'язків між ознаками.
3. Поняття кореляційно-регресійного аналізу.
4. Поняття парної та множинної регресії, їх рівняння.
5. Суть методу найменших квадратів.
6. Використання методу найменших квадратів для визначення параметрів лінійної парної регресії.
7. Коефіцієнт кореляції та його властивості.

7. РЯДИ ДИНАМІКИ

7.1. Елементи та види рядів динаміки

Рядом динаміки, або *динамічним рядом,* називають ряд розміщених у хронологічній послідовності числових даних (статистичних показників), які характеризують величину суспільного явища на даний момент або за певний період часу.

Ряди динаміки складаються з двох *елементів:* рівнів ряду y_i (i=1...n) та часу t_i. *Рівнями ряду* називають числові дані того чи іншого показника ряду динаміки; вони можуть бути виражені в абсолютних, відносних та середніх величинах і задаватися в табличній формі або графічно. *Час ряду* відповідає конкретним моментам або періодам, до яких належать рівні.

За ознакою часу ряди динаміки можуть бути двох *видів:* моментні та інтервальні. *Моментними* називають такі ряди динаміки, рівні яких фіксують стан явища на даний момент часу (дату).

Приклад. Чисельність населення СРСР за переписами населення (табл. 7.1).

Таблиця 7.1 – Чисельність населення СРСР за переписами населення

Дата	На 15 січня 1970р.	На 17 січня 1979р.	На 12 січня 1989р.
Чисельність населення, млн осіб	241,7	262,4	286,7

Інтервальним називають такий ряд, рівні якого характеризують явище за певний період часу. Прикладом інтервального ряду динаміки є виробництво промислової продукції регіону за 2000-2003 рр. (табл. 7.2):

Таблиця 7.2 – Виробництво промислової продукції регіону за 2000-2003 рр.

Рік	2000	2001	2002	2003
Виробництво промислової продукції, млн грн	741	1294	1544	1599

Рівні інтервальних рядів дають підсумкові, результативні показники, які відповідають інтервалу часу, тому їх можна додавати та ділити. При додаванні рівнів ряду знаходять накопичені підсумки.

7.2. Показники рядів динаміки

Для оцінювання властивостей динаміки у статистиці застосовуються взаємопов'язані характеристики, або *аналітичні показники.* Серед них: *абсолютний приріст*, *темп зростання*, *темп приросту та абсолютне значення одного процента приросту*. Розрахунок таких показників ґрунтується на зіставленні рівнів ряду y_i. Якщо базою порівняння є початковий (постійний) рівень ряду y_0, то відповідні показники називаються *базисними* коли ж база порівняння змінна і відповідає попередньому рівню $y_{i\text{-}1}$, то показники називаються *ланцюговими*.

Абсолютний приріст (або *зменшення)* Δ_i відповідає швидкості зміни рівнів ряду і розраховується як різниця рівнів ряду

а) *базисний* $\qquad \Delta_{i0} = y_i - y_0;$ (7.1)

б) *ланцюговий* $\qquad \Delta_i = y_i - y_{i\text{-}1}.\ i=1\ ...n,$ (7.2)

де n – кількість рівнів ряду динаміки.

Ланцюгові та базисні абсолютні прирости пов'язані між собою залежністю (сума ланцюгових приростів дорівнює кінцевому базисному):

$$\sum_{i=1}^{n}\Delta_{i0} = \sum_{i=1}^{n}\left(y_i - y_{i-1}\right) = y_n - y_0. \tag{7.3}$$

Темп зростання K_i характеризує інтенсивність змін рівнів ряду і виражається у відносних величинах числом або у процентах:

а) базисний $\qquad K_{i0} = \dfrac{y_i}{y_0};$ (7.4)

б) ланцюговий $\qquad K_i = \dfrac{y_i}{y_{i-1}}.$ (7.5)

Добуток ланцюгових темпів зростання дорівнює кінцевому базисному

$$K_1 K_2 K_3 \ldots K_n = \prod_{i=1}^{n} K_i = \frac{y_n}{y_0}. \tag{7.6}$$

Темп приросту T_i виражається у процентах і показує, на скільки рівень y_i більший (менший) від рівня, взятого за базу порівняння:

а) базисний $$T_i = \frac{\Delta_{i0}}{y_0} 100\,\% = \frac{y_i - y_0}{y_0} 100\,\%; \qquad (7.7)$$

б) ланцюговий $$T_i = \frac{\Delta_i}{y_{i-1}} 100\,\% = \frac{y_i - y_{i-1}}{y_{i-1}} 100\,\%. \qquad (7.8)$$

Між темпом приросту і темпом зростання існує такий зв'язок:

$$T_i = K_i - 1. \qquad (7.9)$$

Необхідно зазначити, що в динамічних величинах (коефіцієнтів або процентів) безпосередньо порівнювати рівні можна шляхом визначення їх різниці. Ці різниці отримали назву *пунктів зростання*. Їх обчислюють як різницю рівнів базисних коефіцієнтів (процентів) темпів зростання або приросту двох суміжних періодів. На відміну від темпів приросту, які не можна підсумувати та помножити, пункти зростання можна підсумовувати, в результаті чого дістанемо темп приросту відповідного періоду у порівнянні з базисним.

До складу аналітичних показників можуть бути віднесені *коефіцієнти прискорення (уповільнення)* $K_{пр}$, які розраховуються як відношення двох сусідніх темпів зростання K_i та K_{i-1}, визначених ланцюговим способом:

$$K_{пр} = \frac{K_i}{K_{i-1}}. \qquad (7.10)$$

7.3. Середні показники ряду динаміки

Середні показники: середні рівні динамічного ряду; середні з аналітичних показників.

Методи обчислення *середніх рівнів* динамічних рядів залежать від статистичної структури показників.

В *інтервальному ряді* з рівними інтервалами часу застосовують *середню арифметичну просту*, а для нерівних інтервалів – *середню арифметичну зважену*

$$\bar{y} = \frac{\sum y}{n}; \; \bar{y} = \frac{\sum t_i y_i}{\sum t_i}, \qquad (7.11)$$

де t_i – тривалість часу, протягом якого зберігалося значення рівня y_i; n – кількість рівнів ряду.

У *моментних динамічних рядах* з рівними проміжками між датами середній рівень обчислюється за формулою *середньої хронологічної*

$$\bar{y} = \frac{\frac{1}{2} y_1 + y_2 + \ldots + y_{n-1} + \frac{1}{2} y_n}{n-1}. \tag{7.12}$$

Якщо відрізки часу між датами для моментних рядів *різні*, то використовують формулу середньої арифметичної зваженої

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i' t_i}{\sum_{i=1}^{n} t_i}, \tag{7.13}$$

де y_i' – середні рівні окремих інтервалів часу $y_i' = \frac{(y_i + y_{i+1})}{2}$; t_i – тривалість відповідних інтервалів.

Якщо для *моментного ряду* динаміки є дані тільки про початок і кінець періоду, то середній рівень може бути розрахований за формулою

$$\bar{y} = \frac{y_0 + y_n}{2} \tag{7.14}$$

де y_0, y_n – рівні відповідно на початок і кінець періоду.

До *середніх з аналітичних показників* відносяться такі: *середній абсолютний приріст*; *середній темп зростання*; *середній темп приросту*.

Середній абсолютний приріст $\overline{\Delta}$ характеризує середню швидкість зростання (або зменшення) рівнів ряду динаміки. Для моментних та інтервальних рядів середній абсолютний приріст обчислюється за формулою

$$\overline{\Delta} = \frac{\sum_{i=1}^{n} \Delta_i}{n-1} \text{ або } \overline{\Delta} = \frac{y_n - y_0}{n-1}, \tag{7.15}$$

де (n–1) – кількість ланцюгових абсолютних приростів.

Середній темп зростання $\overline{K}$ показує, в скільки разів у середньому кожен даний рівень більший (або менший) від попереднього рівня. Для рядів динаміки з рівними інтервалами середній темп зростання розраховується за формулою середньої геометричної

$$\overline{K} = \sqrt[n]{K_1 K_2 \ldots K_n} \text{ або } \overline{K} = \sqrt[n-1]{\frac{y_n}{y_0}}, \tag{7.16}$$

де K_i – темп зростання за окремі періоди часу; $(n-1)$ – число ланцюгових темпів зростання; y_0, y_n – початковий та кінцевий рівні ряду динаміки.

На основі середнього темпу зростання визначають *середній темп приросту* $\overline{T}$, який показує, на скільки процентів у середньому збільшується (зменшується) цей рівень порівняно з попереднім. Його обчислюють за формулою

$$\overline{T} = \overline{K} - 1. \tag{7.17}$$

Для всебічної характеристики зміни соціально-економічних явищ у часі визначення тільки показників динаміки та їхніх середніх величин не досить. У зв'язку з цим статистика рекомендує ряд спеціальних прийомів обробки та аналізу рядів динаміки.

7.4. Методи обробки динамічних рядів

Вивчаючи ряди динаміки, дослідники намагаються виявити головним чином загальну тенденцію (тренд) у змінах рівнів ряду, тобто основну закономірність розвитку явища, яка вільна від дії різних випадкових факторів. Для цього ряди динаміки підлягають спеціальній обробці – *вирівнюванню*. Вона дозволяє характеризувати особливості зміни за часом динамічного ряду в найбільш загальному вигляді, вважаючи, що через фактор часу можна передати вплив усіх головних факторів.

До *способів і методів вирівнювання* динамічних рядів можуть бути віднесені такі:

а) збільшення інтервалів;

б) обчислення середніх рівнів для збільшених інтервалів;

в) визначення ковзкої середньої;

г) аналітичне вирівнювання.

Найбільш простим способом вирівнювання рядів є *збільшення їх інтервалів.* Суть цього підходу полягає в тому, що первинний ряд динаміки перетворюється і замінюється іншим, рівні якого належать до більших за тривалістю періодів часу (денні інтервали замінюються п'яти- або десятиденними, місячні інтервали – квартальними і т.ін.). Знов утворений ряд буде містити збільшені рівні, які отримані підсумуванням рівнів первинного ряду абсолютних величин. При цьому відхилення в рівнях, обумовлених випадковими причинами, взаємно гасяться, згладжуються і більш ясно виявляються в дії основні фактори зміни рівнів, тобто загальна тенденція.

Розглянемо використання способу збільшення інтервалів за даними реалізації телевізорів у магазинах міста (табл. 7.3):

Таблиця 7.3 – Реалізація телевізорів у магазинах міста

Місяць	Реалізація, шт.	Місяць	Реалізація, шт.
Січень	366	Липень	380
Лютий	310	Серпень	381
Березень	296	Вересень	392
Квітень	380	Жовтень	444
Травень	336	Листопад	382
Червень	295	Грудень	398

Різні напрями змін за окремими місяцями рівнів даного ряду динаміки затрудняє висновки про основну тенденцію продажу телевізорів. Рішення цієї задачі спрощується, якщо відповідні місячні рівні об’єднати у квартальні: I квартал – 972 шт.; II квартал – 1011 шт.; II квартал – 1153 шт.; IV квартал – 1224 шт. Після збільшення інтервалів основна тенденція зростання продажу телевізорів стає явною: 972<1011<1153<1224.

Частковим випадком розглянутого способу є обчислення *середніх рівнів для збільшених інтервалів.* При цьому збільшені рівні ряду динаміки заміняються середніми рівнями збільшених інтервалів.

Одним із поширених простих методів вирівнювання динамічних рядів є їх згладжування за допомогою *ковзної (плинної) середньої.* Суть методу полягає в тому, що для первинного ряду динаміки формуються збільшені інтервали, які складаються з однакової кількості рівнів *m.* Кожен послідовний інтервал отримують послідовним зміщенням від початкового на один рівень. Тоді для нових інтервалів розраховуються середні рівнів $\bar{y}_1 = \dfrac{y_1 + y_2 + \ldots + y_m}{m}$,

$\bar{y}_2 = \dfrac{y_2 + y_3 + \ldots + y_{m+1}}{m}$ і т.д., які як би "згладжують" інтервали і "плинуть по динамічному ряду з кроком, рівним одиниці". Отримуємо новий ряд, зіставлений із ковзних середніх. Кожна із середніх належить до середини укрупненого інтервалу, тому технічно зручніше зіставляти збільшені інтервали із непарної кількості рівнів m (три, п'ять, сім тощо). Обчислення ковзної середньої для парної кількості рівнів є незручним. Це зумовлено тим, що середня може бути віднесена між двома рівнями і тому необхідна додаткова процедура – *центрування*: обчислення середньої із двох суміжних середніх для кожного інтервалу. В результаті новий динамічний ряд, побудований із ковзних середніх, дає виразну тенденцію розвитку явища за рахунок усування коливань рівнів внаслідок випадкових причин.

Недоліком вирівняного ряду методом ковзної середньої є те, що такий ряд "скорочується" порівняно з первинним на $(n – 1)/2$ рівнів ряду з одного та другого кінця (під n мають на увазі кількість рівнів первинного ряду, з яких визначають ковзні середні).

Використання в аналізі рядів динаміки способу збільшення інтервалів та методу ковзної середньої дозволяє виявити тренд для його опису, але отримати узагальнюючу статистичну оцінку тренду цими підходами неможливо. Рішення цієї задачі – вимір тренду – досягається *методом аналітичного вирівнювання.*

Суть аналітичного вирівнювання динамічних рядів полягає в тому, що фактичні рівні ряду замінюються плавними рівнями, обчисленими на основі певної прямої чи кривої, обраної в припущенні, що вона найточніше відображає загальну тенденцію явища.

В основі методу лежить встановлення функціональної залежності рівнів ряду від часу $Y_t = f(t)$ з використанням кореляційно-регресійного аналізу. Практикою статистичних досліджень встановлено, що прийняття тої чи іншої аналітичної функції здійснюється за таких умов, наприклад:

- вирівнювати динамічні ряди за *рівнянням прямої лінії* доцільно тоді, коли більш або менш постійні ланцюгові абсолютні приросту, тобто тоді, коли рівні ряду змінюються приблизно в арифметичній прогресії;

- вирівнювання динамічних рядів за *рівнянням квадратичної параболи* необхідно використовувати у тих випадках, коли зміна рівнів ряду відбувається з приблизно рівномірним прискоренням або уповільненням ланцюгових абсолютних приростів;
- вирівнювання за *степеневою функцією* доцільно використовувати тоді, коли рівні ряду динаміки виявляють тенденцію до сталості ланцюгових темпів зростання, тобто у випадку зміни рівнів ряду динаміки в геометричній прогресії.

Розрахунок параметрів математичних функцій здійснюється *методом найменших квадратів.* Він дає можливість знайти ту залежність, яка найближче проходить до точок фактичних даних на графіку в осях координат "$t – y$", тобто дає найменшу суму квадратів відхилень фактичних значень результативної ознаки y від вирівняних (теоретичних) значень Y_t.

У практичній діяльності може виникнути необхідність інтерполяції або екстраполяції рядів динаміки. Найдосконалішим при цьому є вирівнювання їх за певним аналітичним рівнянням.

Інтерполяція – це знаходження відсутніх проміжних рівнів ряду. Знаючи рівняння тренду для обчислення теоретичних рівнів і підставляючи в нього проміжне значення t між заданими, можна визначити відповідний йому теоретичний рівень результативного фактора Y_t.

Екстраполяція використовується при прогнозуванні суспільних явищ у майбутньому з припущенням, що виявлена тенденція буде зберігатися і надалі за межами досліджуваного ряду динаміки.

7.5. Вимірювання сезонних коливань у рядах динаміки

Ряди динаміки можуть бути представлені у вигляді суми таких *складових:* основної тенденції розвитку – *тренду*; *сезонної* (періодичної) компоненти; *випадкової* компоненти. Методи виділення тренду в динамічних рядах, у яких випадкові фактори зведені до мінімуму, вже були розглянуті. Зупинимося на сезонній складовій динамічного ряду, яка має важливе практичне значення.

Поквартальні або помісячні рівні багатьох показників соціально-економічних явищ суттєво залежать від *сезонності* (сезонних коливань, сезонної хвилі), тобто від більш-менш постійно повторюваних із року в рік коливань рівнів рядів динаміки. У більшості випадків ці коливання пов'язані зі зміною пори року. Такі коливання спостерігаються в багатьох галузях господарювання, наприклад: використання електроенергії залежно від пори року; нерівномірність виробничої діяльності в галузях харчової промисловості, яка пов'язана з переробкою сільськогосподарської сировини; перевезення пасажирів транспортом тощо. Значним коливанням у внутрішньоміській динаміці підлягають грошовий обіг і товарооборот. Найбільші грошові прибутки населення має у III та IV кварталах. Максимальний обсяг роздрібного товарообороту припадає на кінець кожного року. Попит на різні види послуг, виробництво молока, яєць, м'яса, вовни, вилов риби коливаються сезонно.

Сезонні коливання негативно впливають на результати виробничої діяльності та спричиняють порушення ритмічності виробництва. Тому господарчі організації використовують різні заходи для згладжування сезонності за рахунок раціонального об'єднання галузей, механізації трудомістких процесів, утворювання агропромислових фірм тощо.

Комплексне регулювання сезонних змін за окремими галузями економіки повинно основуватися на дослідженнях сезонних коливань.

У статистиці існує ряд *методів* вивчення та виміру сезонних коливань:

а) метод абсолютних різниць;

б) метод відносних різниць;

в) побудова індексів сезонності;

г) побудова аналітичної моделі.

За *методом абсолютних різниць* сезонні коливання характеризуються величинами

$$\Delta_i = \bar{y}_i - \bar{y}_{0i}, \qquad (7.18)$$

де Δ_i – абсолютні відхилення фактичних рівнів або середніх місячних (квартальних) рівнів $\bar{y}_i$ від загальної середньої або трендового *i*-го рівня $\bar{y}_{i0}$.

За *методом відносних різниць* сезонні коливання описуються залежністю

$$\delta_i = \frac{\bar{y}_i - \bar{y}_{0i}}{\bar{y}_{0i}} \qquad (7.19)$$

і можуть бути виражені у відносних величинах або процентах.

Графічне зображення абсолютних або відносних різниць рівнів за місяцями (кварталами) року наочно ілюструють сезонну хвилю.

Якщо значення $\bar{y}_{0i}$ відраховується від вирівнюваного рівня тренда, то для побудови останнього використовується метод ковзної середньої або аналітичне вирівнювання.

Замість відносних різниць за кожен місяць може бути розрахований *індекс сезонності,* який визначається як відношення середнього рівня відповідного місяця до загальної середньої

$$\bar{I}_{S\,i} = \frac{\bar{y}_i}{\bar{y}_{\text{заг.}}} 100\,\%. \qquad (7.20)$$

Індекси сезонності можуть бути розраховані і як відношення фактичного рівня відповідного місяця до рівня, розрахованого за рівнянням тренда.

Сезонна хвиля може бути виділена і при утворенні *аналітичної моделі,* яка основується при дослідженні явищ періодичного типу на використанні соціального типу рівняння – *ряду Фур'є:*

$$Y_t = a_0 + \sum_{i=1}^{k} \left(a_k \cos kt + b_k \sin kt\right), \qquad (7.21)$$

де a_0, a_k, b_k – параметри, які підлягають визначенню; k – кількість членів ряду Фур'є.

Для обчислення параметрів рівняння використовують *метод найменших квадратів*

$$\sum_{i=1}^{k} \left(y_i - Y_t\right)^2 = \min\,. \qquad (7.22)$$

На основі умови (7.22) формують систему нормальних рівнянь, рішення якої дає формули для обчислення параметрів. Аналітичну модель сезонності ряду (7.21) використовують у практичних розрахунках при значеннях k від 2 до 4. При цьому *глибину сезонності* вимірюють за допомогою індексів сезонності.

Загальним *показником* сили коливання динамічного ряду сезонності за рік є *середнє квадратичне відхилення індексів сезонності,* виражене у процентах:

$$\sigma_{Si} = \sqrt{\frac{\sum_{i=1}^{k}\left(I_{Si} - 100\,\%\right)^2}{12}}\,. \qquad (7.23)$$

Чим менша величина цього показника, тим меншою є сезонність досліджуваного явища.

Контрольні запитання

1. Поняття, елементи та види рядів динаміки.
2. Поняття аналітичних показників рядів динаміки та їх види.
3. Суть показника абсолютного приросту, темпу зростання, темпу приросту та коефіцієнтів прискорення.
4. Види середніх показників.
5. Методи обчислення середніх рівнів динамічних рядів.
6. Середні аналітичні показники.
7. Способи і методи вирівнювання рядів динаміки.
8. Суть способу збільшення інтервалів.
9. Суть способу ковзної середньої.
10. Суть методу аналітичного вирівнювання динамічного ряду.
11. Центрування динамічного ряду.
12. Суть інтерполяції та екстраполяції в рядах динаміки.
13. Методи вивчення та виміру сезонних коливань.

8. ІНДЕКСИ ТА ЇХ ВИКОРИСТАННЯ В ЕКОНОМІКО-СТАТИСТИЧНИХ ДОСЛІДЖЕННЯХ

8.1. Поняття індексів та їх значення у статистико-економічному аналізі

Термін "індекс" походить від латинського слова "*index*" і в перекладі означає покажчик, показник. Індекси займають одне з найважливіших місць серед узагальнюючих статистичних показників. З їх допомогою досліджується роль окремих факторів у формуванні економічних показників на макро- і мікрорівні, виявляються резерви виробництва, проводиться зіставлення суспільних явищ у міжнародному масштабі тощо.

Індекс – це відносна величина порівняння, яка характеризує зміну соціально-економічних явищ і процесів у часі, просторі або порівняно з планом (нормою, стандартом).

Формою вираження індексів є коефіцієнти або проценти.

Особливістю індексів є те, що на відміну від інших відносних величин індекси характеризують складові явища, елементи яких не підлягають підсумуванню. *Наприклад*, для товарів з різними споживчими властивостями: молока – в літрах, м'яса – в центнерах тощо. Крім того, індекси завжди характеризують співвідношення однойменних явищ – цін, собівартості, продуктивності праці та ін., що відображається в назві індексів.

За допомогою *індексів* вирішуються такі *основні завдання*:

1) характеристика загальної зміни складного економічного явища у динаміці, територіальному порівнянні, зіставленні з нормативами, планами, прогнозами (наприклад, зміни вартості виробленої продукції, зміни витрат на виробництво, зміни собівартості, продуктивності праці; порівняння споживання продуктів харчування на душу населення в Україні та інших країнах тощо);

2) виявлення у показника складного явища впливу окремих факторів на результативний показник (наприклад, вплив зміни рівня цін і зміни кількості проданих товарів на обсяг товарообороту; виявлення впливу на зростання випуску продукції збільшення чисельності працівників, з одного боку, і збільшення продуктивності праці, з іншого);

3) вивчення динаміки середніх величин та оцінка впливу структурних зрушень на зміну середньої величини (наприклад, оцінка середньої собівар-

тості за групою підприємств з різним рівнем собівартості при випуску однорідної продукції).

Методологія побудови та використання індексів у статистико-економічному аналізі називається *індексним методом*.

Важливою особливістю індексів є те, що їм притаманні синтетичні та аналітичні властивості. *Синтетичні властивості* індексів полягають в тому, що з їх допомогою здійснюється з'єднання (агрегування) в ціле різнорідних одиниць статистичної сукупності. *Аналітичні властивості* індексів проявляються в тому, що за допомогою індексного методу виявляється вплив факторів на зміну досліджуваного показника.

В індексному методі застосовується певна система умовних позначень, за допомогою яких будують і записують індекси. Кожна досліджувана величина має своє позначення у вигляді відповідної літери англійського алфавіту:

а) *кількісні або об'ємні показники:* q – обсяг виготовленої продукції або кількість проданого товару певного виду в натуральному вираженні; T – загальна кількість відпрацьованих людино-годин чи людино-днів (загальні витрати робочого часу на виробництво продукції) або середньооблікова чисельність працівників; h – розмір посівної площі;

б) *якісні показники:* p – ціна одиниці товару чи продукції; z – собівартість одиниці продукції; $t = \frac{T}{q}$ – витрати робочого часу (праці) на виробництво продукції, тобто її трудомісткість; $\overline{q} = \frac{q}{T}$ – середній випуск продукції в розрахунку на одного працівника чи на один людино-день (людино-годину), тобто продуктивність праці; y – врожайність певної культури з 1 га;

в) *показники, що отримані як добуток якісного та кількісного показників:* pq – вартість випуску продукції або загальна вартість проданого товару певного виду (товарооборот); zq – загальна собівартість продукції певного виду, тобто витрати на її виробництво; $tq=T$ – загальні витрати робочого часу на випуск продукції певного виду; yh – валовий збір певної сільськогосподарської культури.

У використанні індексів при динамічних або просторових порівняннях використовують *спеціальні позначення*. Період або об'єкт, з яким порівнюють, називають *базисним*, а період чи об'єкт, який порівнюють, – *поточним*. Дані базисного періоду позначають підрядковим знаком "0", а звітного – "1". *Наприклад*, кількість продукції, виробленої за базисний звітний період, позначають відповідно до q_0 і q_1. Щоб позначити конкретно плановий рівень, пишуть "*пл*"; наприклад, кількість продукції за планом позначають $q_{пл}$. Індекси у вираженні коефіцієнта визначають з точністю 0,0001, що зумовлено взаємозв'язком індексів.

В індексах є дві величини: одну, зміну якої вивчають при використанні індивідуальних та загальних індексів, називають *індексованою*; другу, постійну, у загальних індексах, що приводить різнорідні елементи сукупності до порівнюваного виду, – *сумірником* (для індексів кількісних показників) або *вагою* (для індексів якісних показників).

8.2. Класифікація індексів

У статистичному аналізі використовують різні форми і види індексів, що зумовлює потребу у відповідній їх класифікації. *Індекси можуть бути класифіковані за такими ознаками:*

а) за мірою охоплення елементів сукупності;
б) за базою порівняння;
в) за видом об'єкта порівняння;
г) за видом ваги (сумірника);
д) за формою побудови;
ж) залежно від змісту та характеру індексованої величини;
з) за об'єктом дослідження;
к) за складом явища;
л) за періодом розрахунку.

За мірою охоплення елементів сукупності розрізняють індивідуальні та загальні (зведені) індекси.

Індивідуальні індекси – це відносні показники, які характеризують зміну в динаміці або відображають співвідношення в просторі якогось одного виду одиниць явища. Індивідуальні індекси позначаються буквою *і* та відображають зміну тільки одного елемента сукупності (наприклад, видобуток вугілля на шахті, ціна на картоплю з сільгосппідприємства). Так, i_q –

індивідуальний індекс обсягу продукції, i_p – індивідуальний індекс цін тощо.

Загальні (зведені) індекси позначають буквою I та характеризують динаміку складного явища, елементи якого не піддаються безпосередньому підсумуванню в часі, просторі чи порівняно з планом (наприклад, видобуток вугілля кількома шахтами, зміна цін на картоплю сільгосппідприємствами району). Так, I_q – загальний індекс фізичного обсягу продукції, I_p – загальний індекс цін тощо. У статистичному аналізі використовуються також *групові*, або *субіндекси*, які охоплюють частини цілого (наприклад, індекси продукції за окремими галузями).

За базою порівняння розрізняють базисні та ланцюгові індекси. У *базисних індексах* усі періоди порівнюють з одним, взятим за базу, а в *ланцюгових* – кожен наступний період порівнюють з попереднім.

За видом об'єкта порівняння розрізняють динамічні, територіальні індекси та індекси порівняно з планом (нормою, стандартом). *Динамічні індекси* характеризують зміну явища за часом (ціни, собівартості, продуктивності праці тощо). *Територіальні індекси* відповідають зіставленню показників за відповідними географічними територіями (країнами, регіонами, областями і т.д.). *Індекси порівняно з планом* характеризують стан діяльності підприємств (організацій, установ) на даний поточний період порівняно з установленим планом (стандартом, нормою).

Для загальних (зведених) *індексів за видом ваги* (*сумірника*) розрізняють *індекси з постійними вагами* та *індекси зі змінними вагами*.

За формою побудови, залежно від методології розрахунку, загальні індекси поділяють на агрегатні та середні індекси. *Агрегатні індекси* за рахунок введення сумірника (ваги) в чисельник і знаменник індексу дозволяють здійснити поєднання різнорідних елементів для характеристики складних явищ. *Середні індекси* використовуються у формі середньозважених індексів, коли індексована величина виражається через індивідуальні індекси, а також у формі середніх індексів середніх величин для вивчення динаміки середніх величин.

Залежно від змісту та характеру індексованої величини розрізняють індекси *кількісних* (об'ємних) показників (наприклад, фізичного обсягу продукції) та *індекси якісних показників* (наприклад, цін, собівартості та ін.)

За об'єктом дослідження індекси кількісних показників поділяють на *індекси фізичного обсягу продукції, територіальні індекси, індекси розміру* та *структури посівних площ* тощо. Індекси якісних показників – це індекси цін, собівартості, продуктивності праці та інші.

За складом явища розрізняють індекси постійного та змінного складу, структурних зрушень. Індекси, в яких змінюється одна величина, називають *індексами постійного складу* (індекси цін, собівартості та ін.), а дві і більше величини – *індексами змінного складу* (індекси вартості, обсягу продукції, загальних витрат, валового збору тощо). Відношення індексу змінного складу до індексу постійного складу дає *індекс структурних зрушень*.

Нарешті, *за періодом розрахунку* бувають *річні, квартальні, місячні* та *тижневі індекси*.

8.3. Індивідуальні індекси

Найбільш простим за індексним методом є розрахунок *індивідуальних індексів*. Вони мають відношення до одного елемента явища і не потребують підсумування. Індивідуальні індекси за своєю суттю є *відносними величинами динаміки*, виконання зобов'язань, зіставлення. Розрахунок індексів виконують шляхом обчислення двох індексованих величин у вигляді звичайного дробу: у чисельнику знаходиться величина поточного (звітного) періоду, яка порівнюється, і позначається підрядним знаком "1" (наприклад, кількість виробленої продукції певного виду у поточному періоді q_1, ціна такої продукції – p_1, і т.д.); у знаменнику знаходиться величина базисного періоду, з яким порівнюється величина поточного періоду, і позначається підрядним знаком "0" (наприклад, кількість виробленої продукції певного виду у базисному періоді q_0, ціна такої продукції – p_0 і т.д.).

Прикладами розрахунку індивідуальних індексів є такі:

а) *для кількісних (об'ємних) показників*:

- *індивідуальний індекс фізичного обсягу продукції*

$$i_q = q_1 / q_0 , \tag{8.1}$$

де q_1, q_0 – кількість виробленої продукції певного виду у поточному і базисному періодах;

- *індивідуальний індекс кількості відпрацьованих людино-днів*

$$i_T = T_1 / T_0 , \tag{8.2}$$

де T_1, T_0 – кількість відпрацьованих людино-днів на виробництво продукції у поточному і базисному періодах;

- *індивідуальний індекс розміру посівної площі*

$$i_h = h_1 / h_0 , \tag{8.3}$$

де h_1, h_0 – розміри посівної площ у поточному і базисному періодах;

б) *для якісних показників:*

- *індивідуальний індекс цін на певний вид товару (продукції)*

$$i_p = p_1 / p_0 , \tag{8.4}$$

де p_1, p_0 – ціна одиниці товару в поточному і базисному періодах;

- *індивідуальний індекс собівартості продукції*

$$i_z = z_1 / z_0 , \tag{8.5}$$

де z_1, z_0 – собівартість одиниці продукції в поточному і базисному періодах;

- *індивідуальний індекс продуктивності праці*

$$i_t = t_1 / t_0 , \tag{8.6}$$

де t_1, t_0 – витрати робочого часу (праці) на виробництво одиниці продукції в поточному і базисному періодах;

в) *для показників, які отримані як добуток якісного та кількісного показників*:

- *індивідуальний індекс вартості продукції (товарообороту)*

$$i_{pq} = i_p \cdot i_q ; \tag{8.7}$$

- *індивідуальний індекс загальної собівартості продукції*

$$i_{zq} = i_z \cdot i_q ; \tag{8.8}$$

- *індивідуальний індекс валового збору певного виду сільськогосподарської продукції*

$$i_{yh} = i_y \cdot i_h , \tag{8.9}$$

де $i_y = y_1 / y_0$ – індивідуальний індекс урожайності культури з 1 га.

Індивідуальні індекси можуть розраховуватись у вигляді індексного ряду за декілька періодів. При цьому існує *два способи розрахунку індивідуальних індексів*: ланцюговий і базисний. При *ланцюговому способі розрахунку* за базу порівняння приймається індексована величина сусіднього минулого періоду. При цьому база розрахунку в ряду постійно змінюється. Наприклад, для індексу фізичного обсягу продукції *ланцюгові індекси* за різними періодами розраховуються так:

$$i_{10} = q_1/q_0; \quad i_{21} = q_2/q_1; \quad i_{32} = q_3/q_2 \text{ і т.д.}$$

При *базисному способі розрахунку* за базу приймається незмінна індексована величина якогось одного періоду. Наприклад, для розглянутого випадку *базисні індекси* фізичного обсягу продукції розраховуються так:

$$i_{10} = q_1/q_0; \quad i_{20} = q_2/q_0; \quad i_{30} = q_3/q_0 \text{ і т.д.}$$

Між ланцюговими і базисними індивідуальними індексами існує такий *взаємозв'язок*: добуток ланцюгових індексів дорівнює базисному індексу крайніх періодів. Наприклад, для індексу фізичного обсягу продукції

$$i_{q10} \cdot i_{q21} \cdot i_{q32} = i_{q30}. \tag{8.10}$$

Частка від ділення наступного базисного індексу на попередній дорівнює відповідному ланцюговому індексу

$$i_{q20} : i_{q10} = i_{q21}. \tag{8.11}$$

Тому при наявності ланцюгових індексів можна перейти до базисних, а при наявності базисних – до ланцюгових без прямого розрахунку.

8.4. Агрегатна форма загальних індексів кількісних показників

Найбільш типовим індексом кількісних показників є *індекс фізичного обсягу продукції*. Тому розглянемо його побудову.

У випадку однорідної сукупності для її характеристики можуть бути використані індивідуальні індекси, які не потребують підсумування елементів цієї сукупності.

У випадку неоднорідної сукупності її елементи не підлягають підсумуванню через різну натуральну суть товару та різні одиниці вимірювання (наприклад, такі товари як мед, крупи, картопля, тканини і т.д. в магазині вимірюються у кілограмах, літрах, метрах тощо). Зіставлення загальних фізичних обсягів реалізованих товарів немає сенсу, тому загальний індекс фізичного обсягу продукції як узагальнюючий показник явища не може розраховуватись як:

$$I_q = \frac{\sum q_1}{\sum q_0}.$$

Для цього потрібно привести різні види товарів до порівняльного вигляду, що складає основу методологічної побудови загальних індексів. *Розглянемо суть цієї методології у випадку побудови агрегатної форми загаль-*

них індексів як найбільш розповсюдженої в економічному аналізі досліджуваних явищ (процесів).

Для того, щоб привести різні види товарів до порівнянного виду і здійснювати підсумування різних видів товарів, чисельник і знаменник складного індексу представляють у вигляді *агрегатів*, *тобто поєднання різнорідних елементів*. Кожен з *агрегатів* у чисельнику і знаменнику індексу *являє собою у вигляді суми* (знак $\sum$) добуток *індексованої величини* (для загального індексу фізичного обсягу продукції це кількість вироблених товарів різних видів у поточному q_1 і базисному q_0 періодах) на незмінну величину для видів товарів – *сумірник*.

Для загального індексу фізичного обсягу продукції як сумірник виступають порівнювальні, фіксовані ціни за товарами p_0 на рівні базисного періоду, що дозволяє усунути їх вплив на зміну обсягу продукції. Введення сумірника в агрегати індексу вирішує проблему підсумування, тобто приведення всіх видів товарів (продукції) до єдиного змісту. У випадку індексу фізичного обсягу продукції – це зіставлення агрегатів у вигляді вартості вироблених товарів у періодах порівняння.

Таким чином, загальний індекс, який знаходиться шляхом порівняння результатів складного явища у поточному і базисному періодах за рахунок введення сумірників (ваг), називається *агрегатним*. Спосіб, за допомогою якого складають загальний індекс таким чином, має назву *агрегатного способу*.

Остаточно загальний (зведений) індекс фізичного обсягу продукції в агрегатній формі, або *агрегатний індекс фізичного обсягу продукції*, записується у вигляді

$$I_{pq} = \frac{\sum q_1 p_0}{\sum q_0 p_0}, \qquad (8.12)$$

де q_1, q_0 – кількість вироблених товарів (обсяг продукції) відповідно у поточному (звітному) та базисному періодах; p_0 – незмінна ціна кожного виду товару у базисному періоді; $\sum q_1 p_0$ – умовний показник, який характеризує вартість товарів у поточному періоді за цінами базисного періоду; $\sum q_0 p_0$ – вартість товарів у базисному періоді.

Розрахований за формулою (8.12) *індекс фізичного обсягу продукції* показує, в скільки разів змінився фізичний обсяг продукції або скільки процентів складає його зростання (зниження) у поточному періоді порівняно з базисним періодом.

Так, наприклад, якщо агрегатний індекс фізичного обсягу продукції дорівнює I_q=1,24, або 124 %, то це означає, що загальний випуск продукції в поточному періоді порівняно з базисним періодом зріс у 1,24 раза, або на 24 % (124 – 100 = 24 %). У випадку I_q<1,0 (або 100 %) говорять про зменшення випуску продукції порівняно з базисним періодом.

Різниця чисельника і знаменника індексу (8.12) $\Delta_q = \left(\sum q_1 p_0 - \sum q_0 p_0\right)$ свідчить про абсолютне зростання (Δ_q>0) або абсолютне зменшення (Δ_q<0) вартості випущених товарів у поточному періоді порівняно з базисним періодом у порівнюваних цінах на рівні базисного періоду.

Агрегатні індекси кількісних показників можуть розраховуватись у вигляді індексного ряду за декілька періодів. При цьому використовуються *ланцюгові та базисні способи розрахунку.*

Наведемо приклади ланцюгових і базисних загальних індексів агрегатної форми фізичного обсягу продукції з постійними та змінними вагами (сумірниками) і покажемо їхній взаємозв'язок.

Ланцюгові індекси з постійними вагами:

$$I_{q10} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \quad I_{q21} = \frac{\sum q_2 p_1}{\sum q_1 p_0}; \quad I_{q32} = \frac{\sum q_3 p_0}{\sum q_2 p_0} \text{ і т.д.} \quad (8.13)$$

Ланцюгові індекси зі змінними вагами

$$I_{q10} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \quad I_{q21} = \frac{\sum q_2 p_1}{\sum q_1 p_1}; \quad I_{q32} = \frac{\sum q_3 p_2}{\sum q_2 p_2} \text{ і т.д.} \quad (8.14)$$

Базисні індекси з постійними вагами:

$$I_{q10} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \quad I_{q20} = \frac{\sum q_2 p_0}{\sum q_0 p_0}; \quad I_{q30} = \frac{\sum q_3 p_0}{\sum q_0 p_0} \text{ і т.д.} \quad (8.15)$$

Базисні індекси зі змінними вагами

$$I_{q10} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \quad I_{q20} = \frac{\sum q_2 p_1}{\sum q_0 p_1}; \quad I_{q30} = \frac{\sum q_3 p_2}{\sum q_0 p_2} \text{ і т.д.} \quad (8.16)$$

Між ланцюговими і базисними агрегатними індексами існує такий взаємозв'язок: для індексів з постійними вагами добуток ланцюгових індексів дорівнює базисному індексу крайніх періодів

$$I_{q10} \cdot I_{q21} \cdot I_{q32} = I_{q30}; \tag{8.17}$$

$$\frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum q_2 p_0}{\sum q_1 p_0} \cdot \frac{\sum q_3 p_0}{\sum q_2 p_0} = \frac{\sum q_3 p_0}{\sum q_0 p_0}.$$

Частка від ділення наступного базисного індексу з постійними вагами на попередній дорівнює ланцюговому індексу

$$I_{q20} : I_{q10} = I_{q21}; \tag{8.18}$$

$$\frac{\sum q_2 p_0}{\sum q_0 p_0} : \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum q_2 p_0}{\sum q_1 p_0}.$$

Аналогічно побудованому агрегатному індексу фізичного обсягу продукції (8.12) – (8.18) можуть бути побудовані агрегатні індекси інших кількісних показників, сумірниками і яких виступають якісні показники на рівні базисного періоду.

8.5. Агрегатна форма загальних індексів якісних і змішаних показників

Загальні індекси агрегатної форми якісних показників (цін, собівартості, продуктивності праці тощо) будуються за тією ж методологією, що і агрегатні індекси кількісних показників. Для приведення якісних показників до порівняльного виду утворюються агрегати в чисельнику і знаменнику індексів у вигляді добутку індексованих величин на відповідні постійні ваги кількісних показників. У більшості випадків ваги фіксуються на рівні поточного (звітного) періоду або (в меншій мірі) – на рівні базисного періоду.

Серед агрегатних індексів якісних показників значна роль відводиться *агрегатному індексу цін* I_p, який в більшості випадків використовується у *двох формах*: індексу Пааше та індексу Ласпейреса.

Індекс цін Пааше запропонований в 1874 р. німецьким економістом Г. Пааше. В індексі як вага використовується обсяг продукції певного виду в поточному періоді q_1. Індекс Пааше розраховується за формулою

$$I_p = \frac{\sum q_1 p_1}{\sum q_0 p_1}, \tag{8.19}$$

де p_1, p_0 – індексовані величини цін на певний вид продукції відповідно у поточному та базисному періодах; $\sum q_1 p_1$ – вартість всієї продукції у поточному періоді; $\sum q_0 p_1$ – умовна вартість продукції поточного періоду за порівнюваними цінами базисного періоду.

Індекс цін Пааше характеризує вплив зміни цін на вартість кількості товарів, які реалізовані в поточному періоді.

Розрахований за формулою (8.19) *агрегатний індекс цін Пааше показує*, в скільки разів збільшився (зменшився) у середньому рівень цін на масу товару, що реалізований в поточному періоді, або скільки процентів складає його зростання (зменшення) в поточному періоді порівняно з базисним періодом.

Наприклад, якщо I_p = 0,98, або 98 %, то це означає, що рівень цін на товари, які реалізовано в поточному періоді, в середньому зменшився в 0,98 рази, або на 2 % (100 – 98 = 2 %) порівняно з базисним періодом. У випадку I_p > 1,0 (або 100 %) говорять про збільшення рівня цін у поточному періоді порівняно з базисним періодом.

Різниця чисельника і знаменника (8.19) $\Delta_p = \left(\sum q_1 p_1 - \sum q_0 p_1\right)$ свідчить про абсолютну економію (–) (Δ_p<0) або абсолютну перевитрату (+) (Δ_p>0) грошових коштів покупців у результаті зміни цін на ці товари.

Проте слід зазначити, що вибір ваг при побудові агрегатного індексу цін *не можна вважати обов'язковим у всіх випадках*. У статистиці ряд задач можуть і повинні вирішуватись по-різному залежно від конкретної мети та особливостей дослідження. Проілюструємо це таким *прикладом*. Відомо, що в період економічної кризи різко зростають ціни. В результаті деякі продукти не споживаються населенням, особливо малозабезпеченим. Припустимо, що в умовному базисному періоді у склад споживання входило 30 найменувань продуктів (q_0 = 30), а в поточному періоді – тільки 25 найменувань (q_1 = 25). Очевидно, що в такій ситуації індекс цін, розрахований за q_1, неправільно відобразить зміну цін на ті продукти, які випали із споживання при надмірному зростанні цін. Тому в подібних випадках більш правильно відобразить зміну цін індекс, побудований за продукцією базисного періоду.

У 1864 р. іспанським економістом Е. Ласпейресом запропонований *індекс Ласпейреса*, де як вага використовується обсяг продукції за різновидністю товарів у базисному періоді q_0. Індекс Ласпейреса розраховується за формулою

$$I_p = \frac{\sum q_0 p_1}{\sum q_0 p_0}, \qquad (8.20)$$

де $\sum q_0 p_1$ – вартість продукції у базисному періоді за цінами поточного періоду; $\sum q_0 p_0$ – вартість продукції у базисному періоді.

Індекс Ласпейреса показує вплив зміни цін на вартість кількості товарів, які реалізовано в базисному періоді.

Таким чином, *індекси цін Пааше і Ласпейреса не ідентичні* і для однакових вихідних даних не співпадають, *оскільки мають різний економічний зміст*: індекс Ласпейреса використовують у прогнозуванні обсягу товарообороту у зв'язку з ймовірною зміною цін на товари в майбутньому періоді. Індекс Пааше використовують при вивченні звітних даних, коли ціллю аналізу є якісна оцінка зміни товарообороту в результаті зміни цін у звітному періоді.

Індекс Ласпейреса (*L*) в ряді випадків більше індексу Пааше (*P*). Ця систематична залежність двох індексів відома в статистиці як *ефект Гершенкрона*.

Враховуючи наявну розбіжність між індексами Пааше і Ласпейреса, І. Фішером у міжнародному зіставленні запропонований "ідеальний індекс" (*індекс Фішера*) як середньогеометрична величина з двох вищезгаданих індексів

$$I_p = \sqrt{PL}. \qquad (8.21)$$

Але цей індекс не набув широкого поширення у статистичній практиці країн світу через відсутність у ньому конкретного економічного змісту.

На теперішній час залишається проблема вибору універсальної системи зважування в агрегатних індексах цін. Проте вона компромісно вирішується використанням індексів Пааше чи Ласпейреса для конкретних умов.

В економічному аналізі явищ і процесів використовуються й інші агрегатні індекси якісних показників: собівартості продукції I_z, продуктивність праці I_t, та ін.

Агрегатний індекс собівартості продукції розраховується за формулою

$$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1}, \qquad (8.22)$$

де z_1, z_0 – собівартість одиниці продукції певного виду відповідно у поточному та звітному періодах (індексовані величини); q_1 – кількість виробленої продукції кожного виду у поточному періоді, яка приймається як вага; $\sum z_1 q_1$ – витрати на виробництво продукції поточного періоду; $\sum z_0 q_1$ – умовні витрати на виробництво тієї ж продукції, якщо б собівартість одиниці продукції була на рівні базисного періоду.

Розрахований за формулою (8.22) індекс собівартості показує, в скільки разів зменшився (зріс) у середньому рівень собівартості на продукцію, вироблену у поточному періоді, або скільки процентів складає його зменшення (зростання) в поточному періоді порівняно з базисним.

Якщо із значення індексу собівартості у процентах відрахувати 100 %, то різниця (I_z – І00) покаже, на скільки процентів у середньому зменшився (збільшився) рівень собівартості на продукцію, вироблену у поточному періоді.

Різниця між чисельником і знаменником індексу $\Delta_z = \left(\sum z_1 q_1 - \sum z_0 q_1\right)$ характеризує економію (–) або перевитрати (+) від зміни собівартості одиниці продукції.

Продуктивність праці – це результат конкретної живої праці, ефективність цілеспрямованої діяльності людей у виготовленні продукції протягом відповідного проміжку часу. Виміряється кількістю споживчих вартостей, вироблених в одиницю часу, або кількістю часу, витраченого на одиницю продукції.

Продуктивність праці важлива для успішного вирішення багатьох соціальних та економічних задач. Тільки внаслідок неухильного зростання продуктивності праці можна забезпечити динамічний продуктивний розвиток виробництва, підвищити рівень життя населення.

Агрегатний індекс продуктивності праці за витратами праці на одиницю продукції розраховується за формулою

$$I_t = \frac{\sum t_0 q_1}{\sum t_1 q_1}, \tag{8.23}$$

де $\sum t_0 q_1$ – умовні затрати робочого часу (трудомісткість) на всю продукцію базисного періоду; $\sum t_1 q_1$ – фактичні затрати робочого часу на всю продукцію поточного періоду.

На відміну від наведених вище формул інших агрегатних індексів, в цьому індексі базисна величина індексованого показника (t_0) знаходиться в чисельнику, а поточна величина (t_1) – в знаменнику. Це відбувається тому, що затрати праці на одиницю продукції і продуктивність праці пов'язані оберненою залежністю.

Розрахований за формулою (8.23) індекс продуктивності праці показує, у скільки разів зріс (зменшився) в середньому загальний рівень трудомісткості поточного (звітного) періоду порівняно з базисним.

Якщо із значення індексу продуктивності праці в процентах відрахувати 100 %, то різниця (I_t – 100) покаже, на скільки процентів у середньому зріс (зменшився) на цей час рівень трудомісткості.

Різниця чисельника і знаменника індексу $\Delta_t = \left(\sum t_0 q_1 - \sum t_1 q_1\right)$ показує зростання (Δ_t >0) або зменшення (Δ_t <0) трудомісткості на всю продукцію базисного періоду порівняно з поточним.

Агрегатні індекси якісних показників можуть розраховуватись у вигляді *індексного ряду*. При цьому, як у наведеному прикладі для агрегатного індексу фізичного обсягу продукції, використовуються ланцюговий та базисний спосіб розрахунку для індексів з постійними та змінними вагами.

До *основних агрегатних індексів змішаних показників* можна віднести індекси вартості (товарообороту) товарів I_{pq}, індекси загальної собівартості продукції I_{zq}, індекси загальних витрат робочого часу I_{tq} тощо.

Такі індекси можна подати у вигляді добутку двох індексів, або *системою індексів*, що зручно для аналізу складного явища під впливом певних факторів.

Оскільки *агрегатний індекс вартості товарів (товарообороту)* можна представити як добуток індексу цін I_p (у формі індексу Пааше) та індексу фізичного обсягу продукції I_q

$$I_{pq} = I_p \cdot I_q,$$

то $$I_{pq} = \frac{\sum q_1 p_1}{\sum q_1 p_0} \cdot \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum q_1 p_1}{\sum q_0 p_0}. \qquad (8.24)$$

Як видно із виразу (8.24), цей індекс являє собою відношення вартості товарів поточного (звітного) періоду до вартості товарів базисного періоду. Індекс показує, в скільки разів зросла (зменшилась) вартість товарів (товарооборот) поточного періоду порівняно з базисним, або скільки процентів складає зростання (збільшення) вартості товарів.

Якщо із індексу вартості, вираженому в процентах, відняти 100 %, то різниця (I_{pq} – 100) показує на скільки процентів змінилась вартість товарів у поточному періоді порівняно з базисним.

Різниця чисельника і знаменника формули (8.24)

$$\Delta_{pq} = \left(\sum q_1 p_1 - \sum q_0 p_0\right)$$

показує, на скільки грошових одиниць змінилась вартість товарів у поточному періоді порівняно з базисним.

Якщо відомі два з індексів (8.24), то на підставі цієї залежності можна знайти третій індекс.

Аналогічно *агрегатний індекс загальної собівартості продукції* I_{zq} можна представити як добуток індексу собівартості I_z та індексу фізичного обсягу продукції за собівартістю I_q у вигляді

$$I_{zq} = I_z \cdot I_q$$

$$\text{або } I_{zq} = \frac{\sum q_1 z_1}{\sum q_1 z_0} \cdot \frac{\sum q_1 z_0}{\sum q_0 z_0} = \frac{\sum q_1 z_1}{\sum q_0 z_0}, \quad (8.25)$$

який показує зіставлення витрат на виробництво продукції у поточному і базисному періодах і виражається у коефіцієнтах або процентах.

Агрегатний індекс загальних витрат робочого часу I_{tq} представляється у вигляді добутку індексу продуктивності праці I_t та індексу фізичного обсягу продукції за продуктивністю праці I_q

$$I_{tq} = I_t \cdot I_q$$

$$\text{або } I_{tq} = \frac{\sum q_1 t_1}{\sum q_1 t_0} \cdot \frac{\sum q_1 t_0}{\sum q_0 t_0} = \frac{\sum q_1 t_1}{\sum q_0 t_0}. \quad (8.26)$$

Його величина дає порівняння витрат робочого часу на продукцію різних видів у поточному і базисному періодах.

Індекс валового збору врожаю I_{ys} можна подати як добуток індексу I_y та структури посівних площ I_s

$$I_{ys} = I_y \cdot I_s$$

$$\text{або } I_{ys} = \frac{\sum s_1 y_1}{\sum s_1 y_0} \cdot \frac{\sum s_1 y_0}{\sum s_0 y_0} = \frac{\sum s_1 y_1}{\sum s_0 y_0} \quad (8.27)$$

8.6. Середньозважені індекси

Агрегатний спосіб представлення загальних індексів у статистиці є найбільш поширеним. Разом з тим використовується й інший спосіб розрахунку загальних індексів як середніх із відповідних індивідуальних індексів, або *середньозважених індексів*.

До розрахунку середньозважених індексів звертаються у тих випадках, коли первинна (вихідна) інформація не дозволяє розрахувати загальний агрегатний індекс. Існують *дві форми середньозважених індексів*: *середньоарифметична* та *середньогармонічна*. Як правило, середній арифметичний індекс застосовується при індексуванні кількісних показників (наприклад, фізичного обсягу продукції), а середній гармонічний – при індексуванні якісних показників (наприклад, цін).

До розрахунку *середнього арифметичного індексу* вдаються тоді, коли індексована величина чисельника виражається через індивідуальний індекс. Наприклад, необхідно обчислити загальний індекс фізичного обсягу продукції I_q, коли з вихідних даних відомі індивідуальні індекси фізичного обсягу ($i_q = q_1/q_0$) і вартість продукції кожного виду за базисний період (q_0p_0). Тоді загальний індекс фізичного обсягу можна визначити як середню арифметичну зважену із індивідуальних індексів. Для цього змінимо невідому кількість продукції звітного періоду (q_1) добутком $i_q q_0$ в чисельнику агрегатного індексу (8.12). Тоді загальний індекс фізичного обсягу продукції набуде вигляду

$$I_q = \frac{\sum q_1 p_0}{\sum q_1 p_0} = \frac{\sum i_q q_1 p_0}{\sum q_0 p_0}. \qquad (8.28)$$

Ця формула являє собою середню арифметичну з індивідуальних індексів фізичного обсягу продукції, зважену за вартістю продукції базисного періоду.

Якщо індексована величина виражається через індивідуальний індекс у знаменнику, то індекс має назву *середнього гармонічного індексу*. Наприклад, відомі індивідуальні індекси цін ($i_p = \frac{p_1}{p_0}$) і вартість кожного виду продукції за поточний (звітний) період (q_1p_1), але невідомі дані про ціну за одиницю продукції за базисний період (p_0). Щоб знайти середній гармонічний індекс цін, у знаменнику агрегатного індексу (8.19) ціну базисного періоду (p_0) замінимо рівним їй відношенням $p_0 = \frac{p_1}{i_p}$. Внаслідок цього знаменник агрегатної форми індексу цін (8.19) набуде вигляду $\sum q_1 p_0 = \sum \frac{1}{i_p} q_1 p_1$, а індекс цін матиме вигляд

$$I_p = \frac{\sum q_1 p_1}{\sum \frac{1}{i_p} q_1 p_1}. \qquad (8.29)$$

Ця формула представляє собою середню гармонічну з індивідуальних індексів цін, зважену за обсягом продукції поточного періоду.

8.7. Загальні індекси середніх величин

У статистико-економічному аналізі доводиться порівнювати в динаміці такі узагальнюючі характеристики якісних показників як середня ціна, середня собівартість, середня продуктивність праці тощо. Оскільки на динаміку середньої впливають не тільки зміни усереднювальної ознаки, а й зміни складу розглядуваної сукупності, вплив кожного з цих факторів оцінюється за допомогою *загальних індексів середніх величин.* Такі індекси утворюють індексну систему, яка для якісних показників складається із трьох елементів:

1) індексів змінного складу $I_x^{\text{ЗС}}$;

2) індексів фіксованого складу $I_x^{\text{ФС}}$;

3) індексів структурних зрушень $I_x^{\text{СЗ}}$, де x – вид розглядуваної ознаки (ціна, собівартість, продуктивність праці тощо).

Індекс змінного складу $I_x^{\text{ЗС}}$ показує відносну зміну розглядуваного середнього рівня ознаки в цілому за розрахунок двох факторів – зміни індексованої ознаки і зміни в структурі сукупності

$$I_x^{\text{ЗС}} = \bar{x}_1 : \bar{x}_0 = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}, \qquad (8.30)$$

де $\bar{x}_1, \bar{x}_0$ – середні ознаки відповідно у поточному та базисному (звітному) періодах; f_1, f_0 – ваги ознаки у зіставлюваних періодах.

Індекс фіксованого складу $I_x^{\text{ФС}}$ характеризує зміну середнього рівня за рахунок лише зміни індексованої величини (ваги постійні) при незмінній структурі сукупності

$$I_x^{\text{ФС}} = \bar{x}_1 : \bar{x}_0 = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1}. \tag{8.31}$$

Індекс структурних зрушень $I_x^{\text{СЗ}}$ показує зміну середнього рівня за рахунок змін лише у структурі сукупності при незмінному значенні ознаки

$$I_x^{\text{СЗ}} = \bar{x}_1 : \bar{x}_0 = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}. \tag{8.32}$$

Формули для середніх індексів підпорядковуються принципу зважування, який забезпечує їх пов’язування в *індексну систему*.

$$I_x^{\text{ЗС}} = I_x^{\text{ФС}} \cdot I_x^{\text{СЗ}}. \tag{8.33}$$

З використанням цієї формули за двома відомими індексами можна розрахувати третій.

Контрольні запитання

1. Поняття індексів та їх особливість.
2. Кількісні, якісні та змішані показники.
3. Ознаки класифікації індексів.
4. Поняття індивідуальних та загальних індексів.
5. Поняття ланцюгових та базисних індексів.
6. Приклади загальних індексів агрегатної форми.
7. Особливості використання загальних індексів цін Пааше і Ласпейреса.
8. Поняття середньозважених індексів та їх види.
9. Особливості використання загальних індексів середніх величин.
10. Суть індексу змінного складу.
11. Суть індексу фіксованого складу.
12. Суть індексу структурних зрушень.
13. Зв’язок між загальними середніми індексами.

СПИСОК ЛІТЕРАТУРИ

1. Лугінін О. Є. Статистика : підручник / О. Є. Лугінін, С. В. Білоусова. – К. : Центр навчальної літератури, 2005. – 580 с.
2. Бек В. Л. Теорія статистики : навчальний посібник / В. Л. Бек. – К.: ЦУЛ, 2003. – 288 с.
3. Єріна А. М. Теорія статистики : практикум / А. М. Єріна, З. О. Пальян. – К. : Товариство "Знання", 2006. – 255 с.
4. Шмойлова Р. А. Теория статистики / Р. А. Шмойлова, В. Г. Минашкин, Н. А. Садовникова; под ред. проф. Р. А. Шмойловой. – М. : Финансы и статистика, 1998. – 576 с.
5. Гусаров В. М. Теория статистики : учебн. пособие для вузов / В. М. Гусаров. – М. : Аудит, ЮНИТИ, 1998. – 247 с.

Навчальне видання

БІЛОЦЕРКІВСЬКИЙ Олександр Борисович

ЗАМУЛА Олексій Олександрович

ШИРЯЄВА Наталя Володимирівна

СТАТИСТИКА

Текст лекцій

для студентів спеціальностей 7.050206

"Менеджмент зовнішньоекономічної діяльності"

і 6.030508 "Фінанси"

Роботу до видання рекомендував проф. В. А. Міщенко

Редактор Л. А. Пустовойтова

План 2009 р. п. 115

Підписано до друку . .09. Формат $60\times 84\ ^1\!/_{16}$. Папір офсет.

Друк - ризографія. Гарнітура Таймс. Ум. друк. арк. 4,4. Обл.-вид. арк. 4,8.

Наклад 100 прим. Зам № . Ціна договірна.

Видавничий центр НТУ "ХПІ".

Свідоцтво про державну реєстрацію ДК № 116 від 10.07.2000р.

61002, Харків 2, вул. Фрунзе, 21

Друкарня НТУ "ХПІ". 62002, Харків 2, вул. Фрунзе, 21.